# PREDICTIONS OF THE PROPAGATION SAW TEST: COMPARISONS WITH OTHER INSTABILITY TESTS AT SKIER TESTED SLOPES

Dave Gauthier*[1] and Bruce Jamieson[1,2]

[1]Department of Civil Engineering, [2]Department of Geoscience, University of Calgary, Alberta, Canada

ABSTRACT: Several new fracture propagation field tests have been presented in recent years. These are designed to provide specific information about propagation propensity; however, each of the more common test methods are thought to be demonstrating at least part of the propagation process, and recent research has shown that the new propagation tests perform well at predicting skier-triggered avalanches. But which test performs best under which conditions? To address this question we compared the predictive success of the new Propagation Saw Test (PST) with that of the Compression test (CT), the Rutschblock test (RB), and the Yellow Flags structural instability index (YF) on skier-tested slopes that did and did not release avalanches in the Columbia Mountains of British Columbia, Canada. The results show that, for our dataset, the combined success rate of the PST in predicting stable and unstable conditions was the highest of the group, although it also had a much larger proportion of potentially dangerous 'false stable' results than the other tests. The CT, RB, and YF methods tended to overestimate instability, but often made correct predictions where the PST was incorrect. Overall, the tests usually performed better in combination than on their own, as each provided slightly different instability information.

KEYWORDS: Propagation propensity, field test, instability assessment, snowpack properties, forecasting

## 1. INTRODUCTION

In recent years, several completely new field tests or methods have been developed, and the interpretation of more common methods has been improved. Fracture Character (FC; van Herwijnen and Jamieson, 2007), shear quality (Johnson and Birkeland, 2002), and Release Type (RT; Schweizer, 2002) observations are becoming widely used and accepted additions to the standard Compression test (CT; Jamieson, 1999) and Rutschblock test (RB; Föhn, 1987) results, and the Yellow Flags (YF) structural stability index has improved snow profile interpretation (Schweizer and Jamieson, 2007). Each of these is thought to provide some information about both the ease of initiation and the propagation propensity of the tested slab and weak layer. In 2006, Simenhois and Birkeland presented their 'Extended Column Test' (ECT) and an excellent validation dataset, as a new method specifically designed to investigate propagation propensity in the field. The Propagation Saw Test (PST) was developed at around the same time, by both Swiss and Canadian researchers independently (Gauthier and Jamieson, 2006, 2008a; Sigrist and Schweizer, 2007). Each is well validated individually (e.g. Gauthier and Jamieson, 2008b), but direct comparisons between the refined and new test methods are lacking. However, many researchers are focusing on this very topic. For example, Birkeland and Simenhois (this volume), and Ross and Jamieson (this volume) present side-by-side comparisons of the PST and ECT. In this paper, we compare the predictive success of the PST with that of the Compression test, the Rutschblock test, and the Yellow Flags structural instability index on skier-tested slopes, some that released avalanches and some that did not, in order to assess the relative strengths and weaknesses of each.

*Corresponding author address: Dave Gauthier, Dept. of Civil Engineering, University of Calgary, Calgary, AB, Canada T2N 1N4; tel: 403-220-7479; e-mail: davidmgauthier@gmail.com

2. METHODS AND DATA

Data for this study were collected mainly at sites near two research study areas in the Columbia Mountains of British Columbia, Canada, during the winter of 2007. At each site, we attempted to trigger an avalanche or whumpf by test-skiing a small slope or flat area. Where no whumpf or avalanche occurred, a portion of the skier's track across the slope was excavated and visually assessed for any weak layer damage or fracturing resulting from the passage of the tester. By doing this, each slope test could be classified separately in terms of initiation and propagation of the weak layer fracture. For example, where a slope test resulted in an avalanche, or where a whumpf or remote avalanche was triggered, the slope and tests and observations from that site could be classified as having both initiation and propagation. Where there was no avalanche, but the weak layer was damaged by the passage of the test skier (Figure 1), the tests and the slope were classified as having initiation but no propagation. If no damage to the weak layer was observed, the classification would be no initiation, but unknown propagation – because it is unclear what would have happened if a fracture on the slope had been initiated. This method provides a unique way to isolate the propagation phase of the fracture process, in order to validate the results of the instability or propagation tests, and assess their predictive accuracy.

On each day, following a ski-test as well as initiation and propagation observations, we did several PSTs, alongside Compression tests with Fracture Character observations (CT-FC), and usually one Rutschblock test with score (RB-S) and Release Type (RB-RT) observations. In addition, we recorded a detailed snow profile, from which the Yellow Flags (YF) instability index could be calculated. For each of these tests or methods, a critical result or observation shown to relate to skier triggering is specified, and compared with the prediction made at the same site by the propagation test. Table 1 summarizes the data included in this study, and

*Figure 1. Example of the excavation of a test-skiers tracks where no avalanche resulted. In this case, weak layer damage below the skis indicates that a fracture was initiated, but did not propagate. (ASARC photo)*

the threshold or critical result thought to indicate instability or propagation propensity for each test method. Note that the published critical test results for the CT-FC, RB-S and RB-RT, and YF methods only refer specifically to skier triggering of avalanches, and do not specifically separate the initiation of weak layer fracture and its propagation. The test results indicating instability are used here as a proxy for propagation to compare with the predictions of the PST, since predicting a slab avalanche is predicting both initiation and propagation. Our analysis is designed to determine if the propagation test performs better or worse than other methods in terms of predicting propagation on adjacent slopes (i.e. avalanches and whumpfs).

To compare the performance of each test method, we begin by compiling the results from each test method individually, and populating a 'contingency table' (Table 2) for each based on the predictions of the test method according to their critical or threshold results, and the result of ski-testing the adjacent slope. Four outcomes are possible: the correct prediction of propagation or avalanche; the correct prediction of no propagation or no avalanche; the 'false

*Table 1. Summary of data collected for this study, and the critical results of each test method indicating instability or propagation propensity.*

| Test method | Critical result indicating instability | Sites/Tests All | Sites/Tests Only with confirmed initiation |
|---|---|---|---|
| Propagation Test (PST) | [1]Cut < 50% of column, no fracture arrest | | |
| | - *All column lengths* | 27/184 | 23/170 |
| | - *Only column lengths 90 cm – 110 cm* | 27/75 | 23/62 |
| Compression test (CT) | [2]Fracture character = sudden planar (SP) or sudden collapse (SC) | 27/58 | 23/50 |
| Rutschblock test (RB) | [3]Score ≤ 3 | 23/23 | 19/19 |
| | [3]Release Type = whole block | 23/23 | 19/19 |
| Yellow Flags (YF) | [5]YF ≥ 5 for weak layer | 27/27 | 20/20 |

*Original sources of critical results: [1]Gauthier and Jamieson, 2008b; [2]van Herwijnen and Jamieson,2007; [3]Föhn, 1987; [4]Schweizer, 2002; [5]Schweizer and Jamieson, 2007.*

unstable' or incorrect prediction of propagation; or the 'false stable' or incorrect prediction of no propagation.

Then, for each test method, we calculated the 'True Skill Statistic' (TSS), which allowed for a comparison of the overall predictive success of the test methods (Equation 1).

$$\text{True Skill Statistic (TSS)} = \left(\frac{\text{Correct (YES)}}{\text{False Stable} + \text{Correct (YES)}}\right) - \left(\frac{\text{False Unstable}}{\text{Correct (NO)} + \text{False Unstable}}\right) \qquad (1)$$

The TSS includes each of the two correct prediction categories and both false predictions, in such a way that the correct predictions increase the value of the statistic, and the false predictions reduce it. Values of the TSS can range from +1 for a perfect set of predictions, to -1 if there are no correct predictions. In the case of avalanche forecasting, the false stable carries much higher consequence than the false unstable; however, this weighting of outcomes is not considered in the TSS.

*Table 2. Format and terminology associated with contingency table of predicted versus observed propagation (e.g. whumpf or avalanche).*

| **Propagation:** | | **Observed** YES | **Observed** NO |
|---|---|---|---|
| **Predicted** | **YES** | Correct (YES) | *False unstable* |
| | **NO** | *False stable* | Correct (NO) |

## 3. RESULTS

Table 3 shows that 77% of PSTs performed as part of this study made correct predictions about propagation on 27 skier-tested slopes. In 18% of tests, no propagation was predicted when it was observed; these false stable tests are of particular concern, especially given their frequency and potential consequences. The True Skill Statistic (TSS) for this classification is 0.58.

Table 4 shows that 79% of Fracture Character observations in Compression tests correctly predicted propagation on 27 skier-tested slopes, with a quite low number of false stable predictions. This leads to a TSS of 0.43. While

*Table 3. Contingency table of success of predictions for the PST. TSS = 0.58*

| Propagation: | | Observed YES | Observed NO |
|---|---|---|---|
| Predicted | YES | 79 (42%) | *9 (5%)* |
| | NO | *34 (18%)* | 65 (35%) |

the frequency of observations on slopes without propagation observed was low, the correct and false unstable predictions in these cases are almost equal. Overall, sudden fractures were observed in 77% of Compression tests, compared with 54% in the much larger dataset of van Herwijnen and Jamieson (2007). In addition, 82% of sudden fractures in the current dataset were found on slopes that had avalanched, compared to 46% in their dataset. This means that the current dataset is biased towards sudden fractures and skier triggering, and that only comparisons to other tests from this or similar datasets are appropriate.

Tables 5 and 6 show the predictions of Rutschblock test score and Release Type versus observations of propagation on 23 skier-tested slopes. Correct predictions were found in 69% of cases using the critical range of test score, compared to 70% using Release Type. Rutschblock score had a higher frequency of false stable results (26%) than the Release Type (22%). Overall, a TSS of 0.48 and 0.40 are calculated for the score and Release Type, respectively.

Table 7 summarizes the success of Yellow Flags in predicting propagation. In 75% of the 27 profiles correct predictions were found, with a

*Table 4. Contingency table of success of predictions for the Compression test Fracture Character. Propagation (avalanche or whumpf) is predicted where fracture character was sudden planar or sudden collapse.TSS = 0.43*

| Propagation: | | Observed YES | Observed NO |
|---|---|---|---|
| Predicted | YES | 37 (63%) | *8 (14%)* |
| | NO | *4 (7%)* | 9 (16%) |

*Table 5. Contingency table of success of predictions for the Rutschblock score. Propagation (avalanche) is predicted where the test score was $\leq 3$. TSS = 0.48*

| Propagation: | | Observed YES | Observed NO |
|---|---|---|---|
| Predicted | YES | 10 (43%) | *1 (4%)* |
| | NO | *6 (26%)* | 6 (26%) |

lower number of false stable results (7%) than false unstable (19%); however, in cases where no propagation was observed on adjacent skier-tested slopes, this method had an equal number of correct and false unstable predictions. The TSS calculated from these outcomes is 0.38.

In results presented for each of the methods, all tests from skier-tested slopes were included, regardless of whether or not the initiation of weak layer fracture was confirmed. In Table 8, the TSS of each test is compared with the propagation test for these days, as well as only on days where initiation of weak layer fracture was confirmed (e.g. avalanche or damaged weak layer). For all days with and without observed initiation, the propagation test has the highest TSS, mainly due to the relatively high frequency of correct predictions of the non-occurrence of propagation, and despite the relatively high frequency of false stable results. Each of the other methods has a TSS greater than zero, indicating that in each the correct predictions occurred more often than the incorrect.

When the TSS is recalculated using only days with observed initiation, the TSS for all methods

*Table 6. Contingency table of success of predictions for the Rutschblock Release Type. Propagation (avalanche) is predicted where whole block releases were observed. TSS = 0.40*

| Propagation: | | Observed YES | Observed NO |
|---|---|---|---|
| Predicted | YES | 11 (48%) | *2 (9%)* |
| | NO | *5 (22%)* | 5 (22%) |

*Table 7. Contingency table of success of predictions for the Yellow flags method. Propagation (avalanche) is predicted where there are ≥ 5 yellow flags for the weak layer. TSS = 0.38*

| Propagation: | | Observed YES | Observed NO |
|---|---|---|---|
| **Predicted** | **YES** | 15 (56%) | *5 (19%)* |
| | **NO** | *2 (7%)* | 5 (19%) |

decreases; however, the PST and Rutschblock test Release Type values only decrease slightly (0.02 to 0.05) while the others have a notable decrease of between 0.19 for CT Fracture Character and RB score, and 0.5 for Yellow Flags. These changes reflect only changes in the relative frequency of correct (NO) and false unstable predictions, as the total number of observations of propagation do not change between the datasets. Note that these datasets are smaller, and the number of occurrences in each category are quite low for the RB score and Release Type as well in Yellow Flags. Because of this, a single observation in any contingency could have a large impact on the TSS, and as such, low confidence in their absolute values is implied, although the trends appear relevant.

Also included in table 8 is the TSS calculated for the PST where only tests with approximately 1 m long test columns are included, as recommended by Gauthier and Jamieson (2008a). Note the dramatic increase in the predictive success of the PST when only standard sized columns are included, and the slight increase in TSS where fracture initiation was confirmed (whether or not an avalanche released).

4. DISCUSSION

Compared with adjacent standard stability tests, the propagation test had the best true skill statistic; each of Compression test Fracture Character, Rutschblock test score and Release Type, and Yellow Flags had lower predictive accuracy than the PST overall, using their respective critical ranges of results. Despite this, the standard stability test methods generally had fewer false stable predictions. When only tests on slopes where initiation was observed are compared and therefore the tests must predict only propagation, the TSS of all the standard methods decreased greatly, with only a small decrease in the PST skill. Where only standard sized PST columns are considered, the TSS increases slightly where initiation was confirmed. This implies that the other methods are better at describing initiation propensity rather than propagation propensity, as each had a relative increase in false unstable predictions (i.e. predicted propagation) where propagation did not occur. Conversely, the PST seems to be the best predictor of propagation in this dataset.

Table 9 summarizes the agreement between individual tests on each day. It is clear that where the PST made dangerous false stable

*Table 8. Summary of data collected for this study, and the critical results of each test method indication instability or propagation propensity.*

| Test method | Critical result | TSS values: *All sites* | TSS values: *Only sites with confirmed initiation* |
|---|---|---|---|
| Propagation Test (PST) | Cut < 50% of test column, no fracture arrest | | |
| | *- All column lengths* | 0.58 | 0.56 |
| | *- Only column lengths 90 cm – 110 cm* | *0.77* | *0.80* |
| Compression test (CT) | Fracture character = SP or SC | 0.43 | 0.24 |
| Rutschblock test (RB) | Score ≤ 3 | 0.48 | 0.29 |
| | Release Type = whole block | 0.40 | 0.35 |
| Yellow Flags (YF) | YF ≥ 5 for weak layer | 0.38 | -0.12 |

predictions at the site of whumpfs or avalanches, it was quite unlikely that any of the other methods would make the same prediction error. However, the table also shows that while the Compression test Fracture Character and Yellow Flags often correctly predicted propagation (or avalanche), they also often predicted it where it did not occur (false unstable), suggesting that they may often overestimate the instability. From a decision-making perspective, this is more desirable than the false stable errors, which the PST often made, despite a high proportion of correct predictions of the non-event where the other methods incorrectly predicted propagation. The comparison highlights the idea that each test method has its own strengths and weaknesses, and supports the practice of many professionals to collect information from several sources prior to making a decision about instability.

## 5. CONCLUSIONS

In general, the PST represents a good proxy for propagation on the slope scale, especially when the recommended column size is used. With a simple interpretation rule, the PST had better overall prediction rate (measured by the TSS) than several more common field tests or instability assessment methods. Specifically, the PST performs best in predicting that propagation *will not* occur compared to the other methods, which often overestimate instability, with the caveat that the PST tends to make more false stable predictions than other methods, especially in softer and thinner slabs. One advantage of the PST, however, is that it seems to work best with slab and weak layer combinations that are unsuitable for testing by other methods. For example, where slab thickness is greater than 1.2 m, the Rutschblock and Compression tests are unreliable, as they rely on dynamic energy delivery down through the column to the weak layer to initiate fracture. The PST has been performed with slabs up to 2.85 m with results that predicted propagation according to the interpretation rules. Preparation time and effort appears to be the only limitation in the depth of weak layers that can be tested.

*Table 9. Summary of percentage of tests showing agreement in predictions with the propagation test. The data set is based on the pairing of tests at each site, such that each propagation test is compared with each stability test from the same site. Indicated percentages represent the rate of agreement between the test methods for each contingency of the propagation test. For example, where the PST made a false stable prediction, only 13% of CT-FC also made a false stable prediction.*

| | **Adjacent tests with same prediction** | | | |
|---|---|---|---|---|
| **PST** | **CT-FC** | **RB-S** | **RB-RT** | **YF** |
| **Correct YES** | 95% | 62% | 70% | 88% |
| **Correct NO** | 29% | 76% | 76% | 41% |
| ***False stable*** | **13%** | **12%** | **16%** | **18%** |
| ***False unstable*** | 100% | 62% | 62% | 100% |

## 6. ACKNOWLEDGEMENTS

Special thanks to the ASARC team, the ACS technicians and forecasters at Glacier National Park, and the patrol at Kicking Horse Mountain Resort. For support we are grateful to the, HeliCat Canada Association, Canadian Avalanche Association, Mike Wiegele Helicopter Skiing, Canada West Ski Area Association, Natural Sciences and Engineering Research Council of Canada. This research was partially supported by the Public Safety and Emergency Preparedness Canada Fellowship in Honour of Stuart Nesbitt White.

## 7. REFERENCES

Birkeland, K., and Simenhois, R., 2008 (this volume). The Extended Column Test: Test effectiveness, spatial variability, and comparison with the propagation saw test. Proceedings: 2008 International Snow Science Workshop, Whistler, BC, *In Press*.

Föhn, P., 1987. The rutschblock as a practical tool for slope stability evaluation. IAHS Publication, 162, 195-214.

Gauthier, D. and B. Jamieson. 2008a. Evaluating a prototype field test for fracture and failure propagation in weak snowpack layers. Cold Regions Science and Technology, 51(2-3), 87-97, doi:10.1016/j.coldregions.2007.04.005.

Gauthier, D. and Jamieson, B., 2008b. Fracture propagation propensity in relation to snow slab avalanche release: Validating the Propagation Saw Test, Geophysical Research Letters, 35, L13501, doi:10.1029/2008GL034245

Gauthier, D., and Jamieson, B., 2006. Understanding the propagation of fractures and failures leading to dangerous and destructive snow avalanches: recent developments. Proceedings: 2006 Annual Conference of the Canadian Society for Civil Engineering, First Specialty Conference on Disaster Mitigation, Calgary, AB, 010-1 – 010-9.

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

Johnson, R., and Birkeland, K., 2002. Integrating shear quality into stability test results. Proceedings: 2002 International Snow Science Workshop, Penticton, BC, pp. 514-520.

Ross, C. and Jamieson, B., 2008 (this volume). Comparing fracture propagation tests and tracking propagation trends over time. Proceedings: 2008 International Snow Science Workshop, Whistler, BC, *In Press*.

Schweizer, J., 2002. The Rutschblock test - Procedure and application in Switzerland. The Avalanche Review, 20, 14-15.

Schweizer, J., and Jamieson, B., 2007. A threshold sum approach to stability evaluation of manual snow profiles. Cold Regions Science and Technology, 47, 50-59.

Sigrist, C., and Schweizer, J., 2007. Critical energy release rates of weak snowpack layers determined in field experiments. Geophysical Research Letters, 34, L03502, doi:10.1029/2006GL028576

Simenhois, R., and Birkeland, K., 2006. The extended column test: a field test for fracture initiation and propagation. Proceedings: 2006 International Snow Science Workshop, Telluride, CO, 79-85.

van Herwijnen, A., and Jamieson, B., 2007. Fracture character in compression tests. Cold Regions Science and Technology, 47, 60-68.