# CAN STABILITY TESTS HELP RECREATIONISTS ASSESS THE LOCAL AVALANCHE DANGER?

Bruce Jamieson[1,2], Jürg Schweizer [3], Pascal Haegeli [4], Cam Campbell [1]
[1] Department of Civil Engineering, University of Calgary, Calgary, Alberta, T2N 1N4, Canada
[2] Department of Geology and Geophysics, University of Calgary
[3] WSL Swiss Federal Institute for Snow and Avalanche Research SLF,CH-7260 Davos Dorf, Switzerland
[4] Avisualanche Consulting, Vancouver, BC, Canada

ABSTRACT: In western Canada, various agencies issue public avalanche bulletins three to seven times per week for regions which range from less than 500 $km^2$ to almost 30,000 $km^2$. Sometimes avalanche danger varies substantially within the larger regions. In this study, we assessed whether the results of local rutschblock tests (including whole block releases) and compression tests (including sudden fractures) could help recreationists assess the local avalanche danger. Since “weekend” recreationists cannot reliably select areas of below average stability for their snowpack tests, especially in wind affected areas, we restricted the test sites to sheltered areas at and below treeline where our observers were likely to get the same results as recreationists. Field studies in the Coast, Columbia and Rocky Mountains yielded stability test results and local danger ratings. After a small number of data were filtered to minimize an observation bias, the results of compression tests and rutschblock tests were assessed using ratings of the local avalanche danger. Without considering the danger rating from the regional bulletin, the results of stability tests correlated weakly but significantly with the local avalanche danger. The score from the rutschblock test, with its greater area, correlated better than any of the compression test variables with the local avalanche danger. Various combinations of the regional danger rating and stability test results were assessed in terms of their performance in recognizing when the local avalanche danger was higher than the regional rating. Again the rutschblock results were more predictive than the compression test results. Some simple results of stability tests such as the observation of sudden fractures in compression tests and the release of the entire block in rutschblock tests showed promising results.

KEYWORDS: Snowpack stability tests, avalanche forecasting, avalanche danger, spatial scale.

## 1. INTRODUCTION

During early and mid winter, some recreationists perform stability tests as part of their usual assessment of the avalanche danger in the area in which they are skiing, snowmobiling or snowboarding, and some do not. The question about the value of stability tests has been phrased “To dig or not to dig?”. In an area where a regional bulletin is available, the danger ratings from the bulletin can be used as an initial estimate of the local avalanche danger in the area of the day’s recreation. Hence, the value of stability tests would seem to be less in areas covered by a regional avalanche bulletin. However, in Canada many recreationists travel in areas not covered by regional forecasts (bulletins) or in areas for which the forecast regions are large and the bulletins issued three times per week (Jamieson, Campbell and Jones, 2006, subsequently referred to as JCJ).

Corresponding author:
bruce.jamieson@ucalgary.ca
1-403-220-7479

For a typical day of backcountry snowmobiling, snowboarding or ski touring, recreationists are exposed to avalanche paths within an area of roughly 10 $km^2$. This is the local scale for which recreationists want to know the avalanche danger. They can use

1. the regional avalanche bulletin (if available)
2. various weather and snowpack observations that do not require digging a pit, and optionally
3. snowpack observations, especially stability tests, that do require digging a single pit.

While there are many weather and snow observations relevant to assessing the local avalanche danger (e.g. McClung and Schaerer, 1993, pp. 124-161; Tremper, 2001, pp. 88-170), we focus on stability tests, which are considered Class I data (McClung and Schaerer, 1993, p. 125). It is impractical for those seeking recreation to spend a lot of time on stability tests or any snowpack observations that require digging a pit. We chose to assess the value of stability tests from a single pit, specifically the rutschblock test (Tremper, 2001, pp. 156-158; Greene and others, 2004, pp. 40-42) and the compression test (Canadian Avalanche Association, 2002, pp. 32-34). We considered including snow profiles and assessing them based on Lemons (McCammon and Schweizer, 2002) or Yellow Flags (Jamieson and Schweizer, 2005) but subsequently excluded them because many recreationists do not observe snow profiles and because the level of detail probably varies substantially between those that do.

Given the variability in stability tests on individual slopes (e.g. Campbell, 2004), how can a stability test based on an area ranging from 0.1 $m^2$ for the compression test to 3 $m^2$ for the rutschblock test be indicative of the avalanche danger in an area of 10 $km^2$ (Bloeschl, 1999; Haegeli and McClung, 2004)? At sites selected by experts such tests have been shown to be indicative of the stability on adjacent slopes (e.g. Föhn, 1987; Schweizer and others, 2005). Because of this scale issue, we recognize that the correlations between the results of tests and the local avalanche danger cannot be strong and cannot be as good as they are for the stability of adjacent slopes.

The spatial variability increases in wind affected areas and the potential correlations between stability tests and local avalanche danger must be reduced. We chose to limit our study to treeline (TL) and below treeline (BTL) areas. If we found correlations, then perhaps a study of alpine areas would be worthwhile.

Comparisons between the regional danger rating and the local danger rating are analyzed in JCJ. In this paper, we focus on using the results of stability tests and optionally regional danger ratings to estimate the local avalanche danger.

This study has three objectives

1. To identify which rutschblock and compression test results, if any, can help recreationists assess the local scale avalanche danger
2. In situations where the regional danger rating is available, to evaluate whether stability tests can improve a recreationists' assessment of the local avalanche danger
3. To identify some limitations of rutschblock and compression tests for the assessing the local avalanche danger.

## 2. REGIONAL AND LOCAL DANGER RATINGS

Regional avalanche bulletins in western Canada include danger ratings and several short paragraphs of text. The text typically explains how the weather and snow conditions are contributing to the avalanche danger and discusses the avalanche danger in terms of the terrain. The danger from the regional forecast (or bulletin), $D_{RF}$, is rated as either Low (1), Moderate (2), Considerable (3), High (4) or Extreme (5). While the numbers for danger ratings are used in some European countries, they are currently not included in Canadian bulletins.

In western Canada, forecast regions vary from 100 $km^2$ to almost 30,000 $km^2$ (JCJ). The largest regions are approximately 250 times larger than the smallest region and 2,500 times larger than the scale of a ski tour (approximately 10 $km^2$). The frequency of bulletins ranges from daily to three times per week, adding an issue of the time scale (JCJ).

The local ratings of avalanche danger and field test results for this study are the same as in JCJ. On each observation day in the winters of 2004-05 and 2005-06, field teams of two or three skilled observers traveled on touring skis to a sheltered site at or below treeline. Although avalanche workers in Canada often probe the snowpack to select a uniform representative site before digging a pit, this practice was discouraged to capture the variability inherent in stability tests performed by recreationists. At the site, the team observed a detailed snow profile (which we did not analyze in this study), two or three compression tests and often one or two rutschblock tests. In addition to the compression test score (number of taps) the observers noted the Fracture Character (van Herwijnen and Jamieson, 2005) which is similar to the Shear Quality (Johnson and Birkeland, 2002; Greene and others, 2004, p. 36-37). In addition to the rutschblock score, the observers noted the amount of the block that released (Schweizer and Wiesinger, 2001). The team also made observations of avalanches and other less formal, but often valuable, observations of snow stability while traveling to and from the site. In addition, they had access to weather, snowpack and avalanche observations from the hosting operation and from neighboring avalanche safety programs. Using all available information, a danger rating for the local area and the current day, called the "local nowcast", $D_{LN}$, was selected by consensus. These local danger ratings were recorded for treeline and for below treeline—provided both could be done with confidence. On most days, ratings were recorded for both treeline and below treeline, yielding two cases per observation day.

## 3. OBSERVATIONS

### 3.1 An observation bias in the data?

During the 2004-05 winter of observations, we were occasionally concerned that the stability test

Table 1. Exclusion of cases in which the local nowcast was changed primarily due to the stability test result

| | Cases with compression tests | | | Cases with rutschblock tests | | |
|---|---|---|---|---|---|---|
| winter | total | changed | excluded | total | changed | ex-cluded |
| 2004-05 | 56 | 0* | 0* | 35 | 0* | 0* |
| 2005-06 | 130 | 22 | 10 | 52 | 5 | 2 |
| Total | 186 | 22 | 10 | 87 | 5 | 2 |

* none excluded because local nowcasts were not recorded before the snowpack observations.

Table 2. Cross tabulation of regional and local danger ratings for cases with compression tests

| Local danger rating $D_{LN}$ | Regional danger rating $D_{RF}$ 1 Low | 2 Mod. | 3 Cons. | 4 High | 5 Ext. | Row totals |
|---|---|---|---|---|---|---|
| 1 Low | 30 | 19 | 2 | 2 | 0 | 53 |
| 2 Mod. | 13 | 49 | 17 | 0 | 0 | 79 |
| 3 Cons. | 2 | 5 | 25 | 1 | 0 | 33 |
| 4 High | 0 | 2 | 2 | 4 | 0 | 8 |
| 5 Ext. | 0 | 0 | 0 | 2 | 1 | 3 |
| Column totals | 45 | 75 | 46 | 9 | 1 | 176 |

results might have a strong influence on the local danger rating and therefore could not be used as independent predictors of the local avalanche danger. However, in most cases we were convinced that our local danger rating was based on a wide variety of correlated information and that the stability test results were not dominating the local ratings. To assess the potential bias, in the following winter we rated the local avalanche danger before and after the snowpack observations including the stability tests. If the danger rating changed, observers recorded the reasons for the change. Out of 130 cases with compression tests in the second winter, the local nowcast was changed 22 times (Table 1). In ten of the 130 cases (8%), the change was primarily because of the compression test results. Out of 52 cases with rutschblock tests, the local nowcast was changed five times. In two of the 52 cases (4%), the change was primarily because of the rutschblock results. We excluded the data from the second winter in which the change was primarily due to the specific stability test results. Given this small rate of change caused primarily by the test results in the second winter, we accepted the data from the first winter, acknowledging that we were including a small percentage of biased data (Table 1). After rejecting these biased data, our dataset consisted of 176 cases with compression tests and 85 cases with at least one rutschblock test.

### 3.2 Frequency of the local and regional danger ratings

For cases with compression tests, the frequencies of the local danger rating are cross tabulated against the regional danger ratings in Table 2 and shown in Figure 1. The cases in which the regional danger rating is the same as the local rating are called *hits* (Wilks, 1995, p. 240), and the diagonal of hits in Table 2 is shaded. The cases in which the regional danger rating is higher than the local nowcast are called “Overs”; these lie above and to the right of the shaded diagonal. The cases in which the regional danger rating is lower than the local nowcast are called “Unders” and lie below and to the left of the shaded diagonal.

For the cases with rutschblock tests, the frequencies of the local danger rating are cross tabulated against the regional danger ratings in Table 3.

Table 3. Cross tabulation of regional and local danger ratings for cases with rutschblock tests

| Local danger rating $D_{LN}$ | Regional danger rating $D_{RF}$ 1 Low | 2 Mod. | 3 Cons. | 4 High | 5 Ext. | Row totals |
|---|---|---|---|---|---|---|
| 1 Low | 13 | 7 | 0 | 0 | 0 | 20 |
| 2 Mod. | 7 | 27 | 8 | 0 | 0 | 42 |
| 3 Cons. | 0 | 1 | 15 | 1 | 0 | 17 |
| 4 High | 0 | 1 | 2 | 2 | 0 | 5 |
| 5 Ext. | 0 | 0 | 0 | 0 | 1 | 1 |
| Column totals | 20 | 36 | 25 | 3 | 1 | 85 |

The overall hit rate in the two winters was 62% for cases with compression tests and 68% for cases with rutschblock tests.

If regional danger ratings are interpreted simply, then Unders may contribute to riskier decisions than Overs. In this study we assume the local danger ratings from the nowcasts are unbiased estimates of the local avalanche danger.

The relative frequency of Overs, hits and Unders can be calculated from the difference $\Delta D$ between the regional danger rating $D_{RF}$ and the local danger rating $D_{LN}$

$$\Delta D = D_{RF} - D_{LN} \qquad (1)$$

For Unders $\Delta D < 0$, for hits $\Delta D = 0$, and for Overs, $\Delta D > 0$. The relative frequency of the Unders, hits and Overs for cases with compression tests and for cases with rutschblock tests are shown in Figure 1. Consequently, the higher rate of Overs compared to Unders in Figure 1 indicates a tendency of regional bulletins to be more cautions than our local danger ratings. Unlike with rutschblock tests, there are a few cases with compression tests in which the local danger rating was two or three steps lower than the regional danger rating.

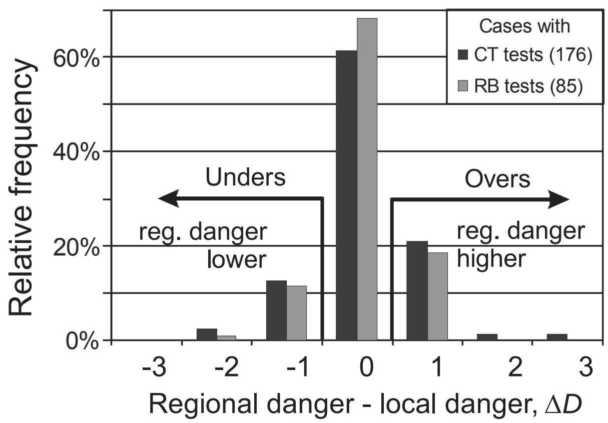


*Figure 1. Relative frequency of difference between regional and local danger rating for cases with compression tests.*

*3.3* *Predictor variables from stability tests*

We analyzed three predictor variables from each set of compression tests at a specific site and five predictors from each set of rutschblock tests (Table 4). In addition to the compression test score CT, i.e. the number of taps for the first fracture, we recorded the number of taps for the first sudden fracture CTS. This allows us to calculate the average number of sudden fractures per compression test nCTS. Observers classified fractures as sudden if they were Sudden Planar (pops) or Sudden Collapse (drops) (van Herwijnen and Jamieson, 2005), or equivalently Shear Quality 1 (Johnson and Birkeland, 2002; Greene and others, 2004). For rutschblock tests, the observers classified the release type as whole block if 90 - 100% of the block released, or most of the block if 50 – 80% of the block released. This is compatible with the Release Type developed by Schweizer and Wiesinger (2001) and Schweizer (2002).

We expect that recreationists with basic training will get the same result as our field team for observations of release type in rutschblock tests or sudden fractures (pops or drops) in compression tests.

## 4. RESULTS AND DISCUSSION

*4.1* *Rank correlations with local danger ratings*

If a variable such as a compression test score or rutschblock score does not correlate with the local avalanche danger then compression or rutschblock tests will not help recreationists assess the local avalanche danger. Accordingly, correlations of the various predictors from Table 4 with the regional and local avalanche danger are shown in Table 5. Significant correlations ($p <$ 0.05) are shown in bold. We used rank correlation because all the predictors are ordered but most lack the interval property. Some of the variables such as the number of whole block releases in compression tests nRBW are only likely to take on a limited number of values such as 0 or 1 and occasionally 0.5. This leads to many ties in the data, particularly for nCTS, nRBW and nRBM. For this reason, we used the gamma correlation in preference to Spearman *R* or Kendall Tau because it explicitly takes ties into account. Gamma $\gamma$ is the difference between the probability that the rank ordering of the two variables agree, minus the probability that they disagree, divided by one minus the probability of ties (Statsoft, 2003).

Table 4. Predictor variables from stability tests

| Variables | **Compression tests** |
|---|---|
| CT | Median of scores (number of taps) from first fracture in each test. If no fracture, CT was set to 35. |
| CTS | Median of scores from first sudden fracture[1] in each test. If no fracture occurred, CTS was set to 35. |
| nCTS | Average number of sudden fractures[1] per compression test. |
| | **Rutschblock tests** |
| RB | Median of first rutschblock score from each test. RB = 7 if there was no planar fracture. |
| RBW | Median rutschblock score of first release of the whole block[2] from each test. RBW = 7 if there was no whole block release. |
| RBM | Median of rutschblock score of first release of the whole block or most of the block[2]. RBM = 7 if there was no release of the whole block or most of the block. |
| nRBW | Average number of whole-block releases[2] per test. |
| nRBM | Average number of releases involving the whole block or most of the block[2] per test. |

[1] Sudden fractures are Shear Quality 1 (Johnson and Birkeland, 2002) or Sudden Planar or Sudden Collapse (van Herwijnen and Jamieson, 2005)

[2] Rutschblock release type (Schweizer and Wiesinger, 2001; Schweizer, 2002)

The correlations in Table 5 are all weak with the highest having an absolute value of 0.39. Strong correlations were not expected because stability test scores vary on the slope scale (e.g. Campbell,

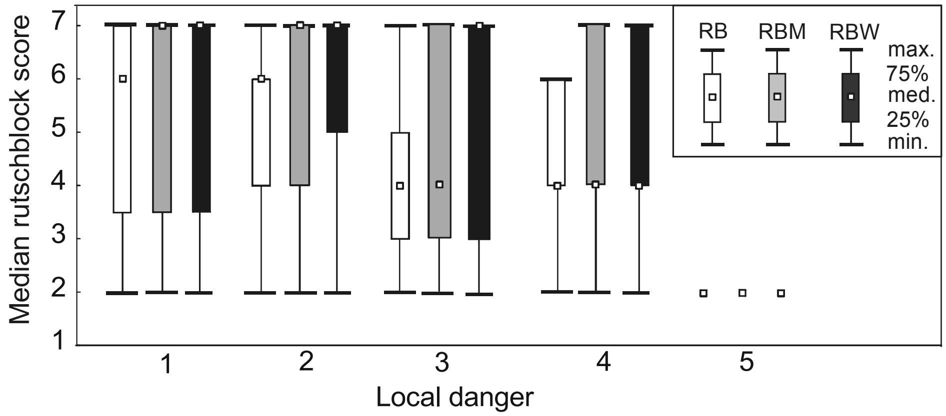

*Figure 3. Rutschblock variables RB, RBW and RBM for each level of local avalanche danger.*

2004) and because the cross sectional area of these compression and rutschblock tests, 0.1 or 3 $m^2$ respectively, is very small in relation to the local and regional scales of avalanche danger. The rutschblock score RB, with its greater area, correlated better than any of the compression test variables at the local scale and at the regional scale (Table 5).

At the local scale, which is considered most relevant for our objectives, CTS correlates with avalanche danger better than CT. This result suggests that observing and classifying the suddenness of the fracture (Johnson and Birkeland, 2002; van Herwijnen and Jamieson, 2005; Greene and others, 2004) can considerably improve the interpretation of test scores from small column tests, as previously shown on the slope scale (van Herwijnen and Jamieson, 2005; Schweizer and others, 2006). The predictors CT and CTS are plotted against the local avalanche danger in Figure 2.

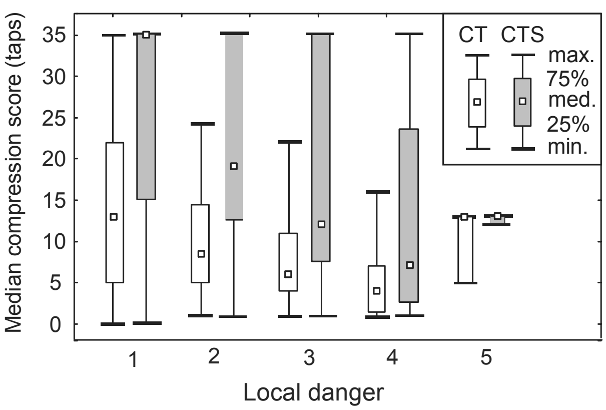

*Figure 2. Compression test variables CT and CTS for each level of local avalanche danger.*

The variables nCTS, nRBW and nRBM correlate significantly with the local avalanche danger. This is of interest since the “suddenness” of the fracture (Johnson and Birkeland, 2002; van Herwijnen and Jamieson, 2005) or the amount of a rutschblock that releases (Schweizer and Wiesinger, 2001; Schweizer, 2002) are observations for which it is reasonable to assume that backcountry recreationists with various levels of training get accurate results.

The sign of the significant correlations is as expected. Lower compression test and rutschblock scores are associated with higher avalanche danger. A higher number of sudden fractures in compression tests or a higher number of whole block or most-of-block releases is associated with higher avalanche danger.

For each rating of local avalanche danger, Figure 3 shows the distribution of the rutschblock variables RB, RBW and RBM. The different correlations for the variables is subtle in Figure 3 although apparent for $D_{LN} = 3$. All the correlations are weak. According to Table 5, RB has a higher gamma correlation with the local danger than RBM or RBW, which is not significantly correlated with the local danger. This is surprising since at the slope scale, the release type supplements the rutschblock score for improved correlations with slope stability (Schweizer and others, 2006).

Although seven of the eight predictors correlate with local avalanche danger in Table 5, only three of eight correlate significantly with the regional avalanche danger. This, combined with the weakness of the correlations, identifies severe limitations of these tests for estimating avalanche danger on the regional scale.

Table 5. Gamma correlations of predictors with local and regional avalanche danger. (Significant correlations ($p < 0.05$) in bold.)

| Danger | nCTS | CT | CTS | nRBW | nRBM | RB | RBW | RBM |
|---|---|---|---|---|---|---|---|---|
| Regional $D_{RF}$ | 0.06 | -0.26 | **-0.19** | 0.08 | 0.16 | **-0.31** | -0.06 | -0.14 |
| Local $D_{LN}$ | **0.23** | **-0.29** | **-0.33** | **0.28** | **0.26** | **-0.39** | -0.23 | **-0.26** |
| $n$ | 176 | 176 | 176 | 85 | 85 | 85 | 85 | 85 |

4.2 _Given the weak correlations with avalanche danger, can the local danger be estimated from stability tests?_

Figures 2 and 3 show that for a given level of avalanche danger, the rutschblock score or range of compression test scores varies widely—too widely for estimating the local danger from rutschblock or compression tests observed at site below treeline or sheltered treeline area. However, experts sometimes interpret the results of surprising low scores as indicating that the avalanche danger is not Low or neither Low nor Moderate. To evaluate this approach, the relative frequency of the rutschblock score RB is tabulated against the maximum local avalanche danger in Table 6, yielding the cumulative frequency distributions by rutschblock score. Since the number of cases is small for some cells in the table, the rutschblock scores are grouped into 2 ≤ RB ≤ 4, 5 or 6, and 7 to smooth the cumulative frequency distributions in Figure 4. In our data, rutschblock scores of 6 or less occurred less than 20% of the time when the danger was Low. Data such as these could be used to develop guidelines for recreationists, e.g. when the rutschblock score is 6 or less, there is only a 15-20% chance that the avalanche danger at or below treeline is Low, or alternatively, there is at least a 75% chance that the danger at or below treeline is Moderate or higher. Such guidelines might help recreationists recognize higher than expected avalanche danger.

Table 6. Relative frequency of the maximum local avalanche danger by rutschblock score

| | Maximum local avalanche danger | | | | | No. of cases |
|---|---|---|---|---|---|---|
| RB | 1 | 2 | 3 | 4 | 5 | |
| 2 | 0.36 | 0.64 | 0.82 | 0.91 | 1 | 11 |
| 3 | 0.14 | 0.43 | 1 | 1 | 1 | 7 |
| 4 | 0.06 | 0.53 | 0.88 | 1 | 1 | 17 |
| 5 | 0 | 0.71 | 1 | 1 | 1 | 7 |
| 6 | 0.18 | 0.89 | 0.93 | 1 | 1 | 28 |
| 7 | 0.60 | 0.87 | 1 | 1 | 1 | 15 |

In Table 5, CTS correlated better than other compression test variables with the local avalanche danger. For cases with compression tests, Table 7 and Figure 5 follow the approach used in Table 6 and Figure 4 for cases with rutschblock tests. Table 7 shows that when CTS ≤ 20, i.e. a sudden fracture occurred within the first twenty taps (average of 2-3 tests), the local danger was Low in less than 20% of cases. In contrast when the first sudden fracture occurred between the 21$^{st}$ and 30$^{th}$ tap (average of 2-3 tests), the avalanche danger was Low in 42% of cases. This suggests that the expectation of Low avalanche danger could be questioned by a sudden fracture within the first twenty taps (average of two or three compression tests).

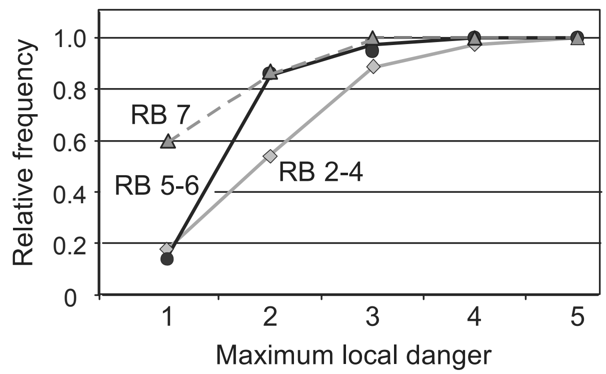


_Figure 4. Relative frequency of rutschblock scores by the maximum local avalanche danger._

Although compression and rutschblock scores have been correlated with stability in adjacent slopes (Föhn, 1987; Schweizer and others, 2005), Figures 4 and 5 show that, in many situations, stability tests from a single pit are—by themselves—poor predictors of the local avalanche danger. This is why experts rely on a wide variety of observations of weather, snowpack and avalanches. In most situations, however, our data support the advice of avalanche experts that stability tests from a single pit are not a sound basis for estimating the local avalanche danger. Systematic approaches, perhaps based on a threshold sum (e.g. McCammon, 2004; Schweizer and others, 2006), that integrate many observations might be developed for local scale decisions.

Table 7. Relative frequency of the maximum local avalanche danger by compression test score CTS

| | Maximum local danger | | | | | No. of cases |
|---|---|---|---|---|---|---|
| CTS | 1 | 2 | 3 | 4 | 5 | |
| CTS ? 10 | 0.19 | 0.38 | 0.30 | 0.14 | 0.00 | 37 |
| 10 < CTS ? 20 | 0.17 | 0.52 | 0.24 | 0.02 | 0.06 | 54 |
| 20 < CTS ? 30 | 0.42 | 0.58 | 0.00 | 0.00 | 0.00 | 24 |
| No sud. fract. | 0.44 | 0.38 | 0.15 | 0.03 | 0.00 | 61 |

4.3 _Is the regional danger rating better than local snowpack observations for estimating the local danger?_

For cases with compression tests, the danger rating from the regional forecast correlates better with the local avalanche danger (Spearman $R$ = 0.61, gamma $\gamma$ = 0.76, $n$ = 176) than any of the stability test variables in Table 5. The rate of agreement, or hit rate (Wilks, 1995, p. 240) between the danger rating from the local nowcast and the regional forecast is 62% for cases with compression tests and 68% for cases with rutschblock tests. So, given the constraints of this study and including our attempt to select sites similarly to recreationists, the regional danger rating is much better than local snowpack tests for estimating the local avalanche danger.

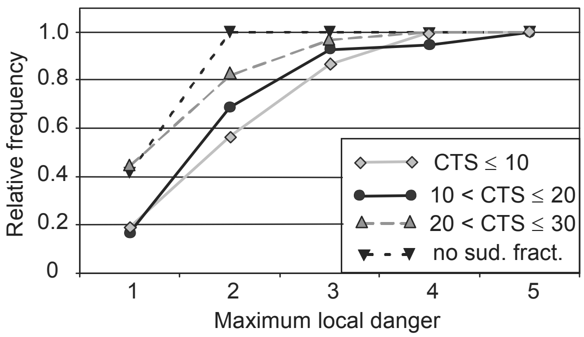


*Figure 5. Relative frequency of compression test variable CTS by the maximum avalanche danger.*

*4.4* *<u>In areas where the regional bulletin is available, can local stability tests help recreationists assess the local avalanche danger?</u>*

In other words, when traveling in area with a regional bulletin, can stability tests help recreationists assess the local avalanche danger? Since a lot of recreation takes place in areas with a regional bulletin, this is a central question of this study.

Table 8. Gamma correlations of predictors with $\Delta D$
Significant correlations ($p < 0.05$) in bold.

| Predictor | $\Delta D$ | $p$ |
|---|---|---|
| nCTs | **-0.24** | **0.001** |
| CT | -0.02 | 0.74 |
| CTs | **0.18** | **0.01** |
| nRBW | **-0.32** | **0.02** |
| nRBM | -0.12 | 0.40 |
| RB | 0.09 | 0.46 |
| RBW | **0.26** | **0.046** |
| RBM | 0.15 | 0.21 |

Table 9. Relative frequency of Unders and Overs by regional danger rating

| | Regional danger rating $D_{RF}$ | | | | |
|---|---|---|---|---|---|
| | 1 Low | 2 Mod. | 3 Cons. | 4 High | 5 Ext. |
| Cases with compression tests | | | | | |
| Unders | 33% (15/45) | 9% (7/75) | 4% (2/46) | 22% (2/9) | 0% (0/1) |
| Overs | 0% (0/45) | 25% (19/75) | 44% (41/46) | 33% (3/9) | 0% (0/1) |
| Cases with rutschblock tests | | | | | |
| Unders | 35% (7/20) | 6% (2/36) | 8% (2/25) | 0% (0/3) | 0% (0/1) |
| Overs | 0% (0/20) | 19% (7/36) | 32% (8/25) | 33% (1/3) | 0% (0/1) |

To assess the potential of combining the regional danger rating with the results of stability tests, Table 8 shows the gamma correlations of the predictor variables with the difference between the regional and local avalanche danger. Four of the predictors, two based on the compression test and two based on the rutschblock test, are significant ($p < 0.05$). Notably, all of the four significant predictors include either the appearance of the fracture in compression tests or the release type in rutschblock tests, both of which Schweizer and others (2006) argue are indicative of fracture propagation.

As a practical example of combining the regional danger rating with stability test results, experts might consider a whole block release in a rutschblock test to be an important indication of local avalanche danger when the regional danger is Low or Moderate but such a result might not be surprising when the regional danger is High.

The following analysis focuses on recognizing Unders since it is particularly important to recognize when the local avalanche danger is higher than the regional danger. Table 9 shows that, in our dataset, Unders (regional danger less than local danger) are more common when the avalanche danger is Low.

To assess the combination of regional danger rating and certain stability test results, we chose to explore the available data with an if-then rule that is similar to the experienced based approach described above:

If <regional condition> and <local stability test condition> then <conclusion about local danger>

The <regional condition> can be of the form $D_{RF} \leq D_{RF}^*$ where $D_{RF}^*$ is some specified threshold of avalanche danger, e.g. Moderate, and <local stability test condition> can be of the form RB ≤ RB* where RB* is a specific threshold rutschblock score, e.g. 3. The <conclusion about local danger> could be qualitative like “be extra cautious” or quantitative like “local danger rating is regional danger rating + 1”. We rejected the quantitative conclusions because we doubt that recreationists quantify extra caution in terms of one or two steps of the danger rating and because there were too few differences of $\Delta D$ = -2 or -3 to assess rules involving such conclusions. Since an Under is exactly the situation in which extra caution may be appropriate, we assessed each rules ability to recognize Unders using the contingency table shown in Table 10.

Table 10. Two by two contingency table for recognizing Unders

| | Observed | | Row totals |
|---|---|---|---|
| Predicted | Unders | Hits + Overs | |
| Unders | a | b | a + b |
| Hits + Overs | c | d | c + d |
| Column totals | a + c | b + d | n = a + b + c + d |

The effectiveness of various rules and the thresholds on the stability test results for recognizing Unders were assessed with the Threat Score TS, the False Alarm Rate FAR and the True Skill Score TSS (Wilks, 1995, p. 240-250) defined as follows:

$$TS = \frac{a}{a+b+c} \tag{2}$$

$$FAR = \frac{b}{a+b} \tag{3}$$

$$TSS = \frac{ad - bc}{(a+c)(b+d)} \tag{4}$$

The Threat Score is the number of times an Under is correctly predicted divided by the number of times a Under was predicted and/or observed. This is an improvement score that can range from 0 when no Unders are correctly predicted (*a* = 0) to 1 when all Unders are correctly predicted and none are incorrectly predicted (*b* + *c* = 0).

The False Alarm Rate is the proportion of predicted Unders that were not observed. The best FAR is 0 (*b* = 0) and the worst value is 1 when no Unders are correctly predicted (*a* = 0).

The True Skill Score or Hanssen-Kuipers discriminant is a measure of the improvement over a random forecast (Wilks, 1995, p. 249) and ranges from negative values for predictions that are worse than random to 1 for perfect predictions.

With the chosen set of stability test results as predictors, we varied the thresholds on the regional danger rating $D_{RF}$* and the threshold on the stability test results until the Threat Score was maximized. In almost all case the True Skill Score was simultaneously maximized. The results for the compression test and rutschblock tests predictors are summarized in Tables 11 and 12, respectively.

Each condition in Tables 11 and 12 represents a way of recognizing Unders. For each condition, TS was maximized when the regional avalanche danger was Low ($D_{RF}$* = 1), probably because Unders occurred most often when the regional danger was rated Low (Table 9).

Table 11. Performance of compression test conditions for recognizing Unders

| Condition for recognizing Unders | | | | | | |
|---|---|---|---|---|---|---|
| If $D_{RF} \leq D_{RF}$* and CTS ≤ CTS* | $D_{RF}$* | CTS* | TS | FAR | TSS | *a* + *b* |
| | 1 | 23 | 0.275 | 0.560 | 0.330 | 25 |
| If $D_{RF} \leq D_{RF}$* and CT ≤ CT* | $D_{RF}$* | CT* | TS | FAR | TSS | *a* + *b* |
| | 1 | 19 | 0.333 | 0.533 | 0.432 | 30 |
| If $D_{RF} \leq D_{RF}$* and nCTS ≥ nCTS* | $D_{RF}$* | nCTS* | TS | FAR | TSS | *a* + *b* |
| | 1 | 1 | 0.275 | 0.560 | 0.330 | 25 |

Table 12. Performance of rutschblock test conditions for recognizing Unders

| Condition for recognizing Unders | | | | | | |
|---|---|---|---|---|---|---|
| If $D_{RF} \leq D_{RF}$* and nRBW ≥ nRBW* | $D_{RF}$* | nRBW* | TS | FAR | TSS | *a* + *b* |
| | 1 | 1 | 0.308 | 0.333 | 0.337 | 6 |
| If $D_{RF} \leq D_{RF}$* and nRBM ≥ nRBM* | $D_{RF}$* | nRBM* | TS | FAR | TSS | *a* + *b* |
| | 1 | 0[a] | 0.292 | 0.650 | 0.465 | 20 |
| If $D_{RF} \leq D_{RF}$* and RB ≤ RB* | $D_{RF}$* | RB* | TS | FAR | TSS | *a* + *b* |
| | 1 | 6 | 0.438 | 0.417 | 0.571 | 12 |
| If $D_{RF} \leq D_{RF}$* and RBW ≤ RBW* | $D_{RF}$* | RBW* | TS | FAR | TSS | *a* + *b* |
| | 1 | 6 | 0.385 | 0.286 | 0.428 | 7 |
| If $D_{RF} \leq D_{RF}$* and RBM ≤ RBM* | $D_{RF}$* | RBM* | TS | FAR | TSS | *a* + *b* |
| | 1 | 5 | 0.385 | 0.286 | 0.428 | 7 |

[a] 7 of the 20 predicted Unders (regardless of nRBM value) occurred in the 11 cases when Unders were observed.
[b] 9 of the 54 predicted Unders (regardless of nRBM value) occurred in the 11 cases when Unders were observed.

In terms of TS or TSS, the conditions based on rutschblock predictors performed better than the conditions based on compression test predictors. For compression test predictors, CT performed better (TS = 0.33, TSS = 0.43) than CTS or nCTS but with a very high False Alarm Rate of 0.53. For rutschblock predictors, RB performed better (TS = 0.43, TSS = 0.57) than RBW, RBM, nRBW or nRBM but with a high False Alarm Rate of 0.42. RBW performed as well as RBM (TS = 0.39, TSS = 0.49, FAR = 0.29). RBW and RBM show promise because they exhibit substantially lower False Alarm Rates than RB and because their values of TS and TSS are only slightly lower than for RB. In our dataset, there were few cases of most-of-block releases and consequently there is no advantage of RBW over RBM. However, we note that Schweizer and others (2006) found that whole block releases correlated much better than most-of-block releases with skier triggered slab avalanches on adjacent slopes (i.e. slope scale).

The optimal threshold for RBW is 6, which includes all whole block releases since there is no release for RB = 7. Thus it seems that performance of RBW ≤ 6 and nRBW ≥ 1should be equal. However, the different performance results from RBW being a median score and nRBW being an average. Also, to simplify the interpretation, we did not try fractional values of the thresholds when optimizing the conditions.

In summary, our method of optimization identifies the potential of stability tests and, in particular of whole block releases and of the rutschblock score, for supplementing the regional danger rating.

## 5. CONCLUSIONS

A large dataset consisting of local danger ratings for areas of approximately 10 $km^2$ and stability tests at and below treeline from seven forecast regions in western Canada were analyzed. After filtering out cases in which the stability test result primarily influenced the local danger rating, there were 85 cases with one or more adjacent rutschblock tests and 176 cases with three adjacent compression tests. Since local danger ratings for both treeline and for below treeline were typically associated with one set of stability tests, the overall number of data points for the study is roughly twice number of sets of compression tests and sets of rutschblock tests.

The danger rating from the regional forecast was by far the best predictor of the local danger since correlations between stability tests and the local danger were consistently weak. Seven of the eight predictors (stability test variables) correlated significantly with local avalanche danger, whereas only three of eight correlated significantly with the regional avalanche danger, identifying severe limitations for the regional interpretation of tests results from a single pit. The rutschblock score RB correlated better than any of the compression test variables at the local scale and at the regional scale.

On the local scale, which was most relevant for our objectives, the compression tests score for the first sudden fracture correlated more strongly with the local avalanche danger than the compression test score for the first fracture (sudden or not), suggesting that observing and classifying the appearance of the fracture (Johnson and Birkeland, 2002; van Herwijnen and Jamieson, 2005; Greene and others, 2004) can considerably add to the interpretation of the test score from small column tests at the local scale.

Observations of sudden fractures in compression tests (independent of score) and of whole block releases in rutschblock tests (independent of score) correlated significantly with the local avalanche danger. This is of interest since the “suddenness” of the fracture or the amount of a rutschblock that releases are practical observations for backcountry recreationists with basic training.

Various conditions for recognizing when the local avalanche danger is higher than the regional danger were assessed. This situation occurred most often when the regional danger was Low and, accordingly, each of the performance measures for the rules were consistently optimized for Low avalanche danger. Rutschblock variables outperformed compression test variables. In terms of the True Skill Sore or Threat Score, the traditional rutschblock score performed best; however, it predicted increased local danger in many cases in which the local danger was not higher than the regional danger. The rutschblock score for the first release of a whole block did not overestimate the local avalanche danger as often and recognized many cases when the local avalanche danger was higher than the regional danger. More data are required before data-based rules or guidelines for interpreting local snowpack observations in conjunction with the regional avalanche danger can be recommended.

Stability tests comprise only a few of the many weather and snowpack observations relevant to assessing the local avalanche danger. This study did not compare the value of stability tests to the numerous other observations, many of which are easier and faster to observe. Also, this study did not assess stability tests in wind affected alpine areas.

## ACKNOWLEDGEMENTS

For meticulous snowpack observations, we thank Paul Langevin, Antonia Zeidler, Ken Matheson, Ali Haeri, Laura Bakermans, James Floyer, Dave Gauthier and Catherine Brown, as well as Gord Irwin, Percy Woods, Brad White, Lisa Paulson, Marc Ledwidge, Steve Michelle, Mike Henderson and Christine Aikens from Banff, Yoho and Kootenay National Parks. The study relied on the logistical support and advice from the Avalanche Control Section of Glacier National Park, BC Ministry of Transportation at Pemberton and Hope, Whistler Blackcomb Ski Resort, Cypress Mountain, Banff, Yoho and Kootenay

National Parks, Kicking Horse Mountain Resort and from Mike Wiegele Helicopter Skiing. Thanks to Susan Hairsine and Clair Israelson for encouragement and to the National Search and Rescue Secretariat for funding for the field work part of this study under the Canadian Avalanche Association's ADFAR project.

REFERENCES

Campbell, C. 2004. Spatial variability of slab stability and fracture properties in avalanche start zones. MSc thesis. Dept. of Civil Engineering, University of Calgary, Calgary, Alberta, Canada.

Canadian Avalanche Association. 2002. *Observation Guidelines and Recording Standards for Weather, Snowpack and Avalanches*. Canadian Avalanche Association, Revelstoke, B.C.

Bloeschl, G. 1999. Scale issues in snow hydrology. Hydrological Processes, 13(14-15), 2149-2175.

Föhn, P.M.B. 1987. The rutschblock as a practical tool for slope stability evaluation. In: B. Salm and H. Gubler, eds., Avalanche Formation, Movement and Effects. International Association of Hydrological Sciences, Publication No. 162, 223-228.

Greene, E.M., K.W. Birkeland, K. Elder, G. Johnson, C. Landry, I. McCammon, M. Moore, D. Sharaf, C. Sterbenz, B. Tremper, and K. Williams. 2004. *Snow, weather, and avalanches: Observational guidelines for avalanche programs in the United States.* American Avalanche Association, Pagosa Springs, Colorado, 140 pp.

Haegeli, P., and McClung, D.M., 2004. Hierarchy theory as a conceptual framework for scale issues in avalanche forecast modeling. Annals of Glaciology, 38, 209-214.

van Herwijnen, A.F.G. and B. Jamieson. 2005. Fracture character in compression tests. Proceedings of the 2004 International Snow Science Workshop in Jackson Hole, Wyoming. USDA Forest Service, Fort Collins, CO, 182-191.

Jamieson, B., C. Campbell, A. Jones. 2006. Spatial and time scale effects in Canadian avalanche bulletins. Proceedings of the 2006 International Snow Science Workshop in Telluride, CO, ISSW 2006 Workshop Committee.

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

Johnson, R. F., and K. W. Birkeland. 2002. Integrating shear quality into stability test results, in Proceedings ISSW 2002, International Snow Science Workshop, edited by J. R. Stevens, pp. 508-513, 29 September-4 October 2002. BC Ministry of Transportation, Victoria, BC, Canada.

McCammon, I. 2004. Evidence of heuristic traps in recreational avalanche accidents: Evidence and implications. Avalanche News 68, 42-50.

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.

McClung, D.M. and P. Schaerer. 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., K. Kronholm and T. Wiesinger. 2003. Verification of regional snowpack stability and avalanche danger. Cold Regions Science and Technology 37(3), 277-288.

Schweizer, J., and T. Wiesinger. 2001. Snow profile interpretation for stability evaluation. Cold Regions Science and Technology 33(2-3), 179-188.

Schweizer, J., C. Fierz and B. Jamieson. 2005. Assessing the probability of skier triggering from snow layer properties. Proceedings of the International Snow Science Workshop in Jackson Hole, Wyoming. USDA Forest Service, Fort Collins, CO, 192-198.

Schweizer, J., I. McCammon and B. Jamieson. 2006. Snow slope stability evaluation using concepts of fracture mechanics Proceedings of the 2006 International Snow Science Workshop in Telluride, CO, ISSW 2006 Workshop Committee.

Statsoft, 2003. Statistica Electronic Manual. Statsoft Inc., Tulsa, OK.

Tremper, B. 2001. Staying Alive in Avalanche Terrain. The Mountaineers Books, Seattle, WA.

Wilks, D.S. 1995. Statistical Methods in the Atmospheric Sciences. Academic Press, San Diego.