Cold Regions Science and Technology 120 (2015) 191–196

Contents lists available at ScienceDirect

Cold Regions Science and Technology

journal homepage: www.elsevier.com/locate/coldregions

Using 2 m Extended Column Tests to assess slope stability

Edward H. Bair [a,b,*], Ron Simenhois [c], Alec van Herwijnen [d], Karl W. Birkeland [e]

[a] Earth Research Institute, University of California, Santa Barbara, CA, USA
[b] US Army Corps of Engineers Cold Regions Research and Engineering Laboratory, Hanover, NH, USA
[c] Coeur, Juneau, AK, USA
[d] WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland
[e] US Forest Service National Avalanche Center, Bozeman, MT, USA

ARTICLE INFO

Article history:
Received 19 November 2014
Received in revised form 24 June 2015
Accepted 26 June 2015
Available online 6 July 2015

Keywords:
Snow
Avalanche
Stability tests
Extended Column Tests

ABSTRACT

Extended Column Tests (ECTs) are used to assess crack initiation and propagation. Previous research shows that tests 90 cm in length may propagate, suggesting instability, while tests 2 m in length may not propagate, suggesting stability, for identical snowpacks. A practical question is: are 90 cm ECTs optimal for assessing stability? To test the added value of 2 m ECTs for stability evaluation, we collected data on 220 ECTs, with 136 side-by-side standard length ECTPs (full propagation indicating instability) followed by 2 m ECTs. We only performed 2 m ECTs after a standard ECT propagated because we assumed 2 m ECTs would not propagate if standard length tests did not. These tests were preceded by an a priori stability assessment. Our results show imbalances for both tests. The ECT had a similar probability of detection (0.88–0.92, POD), i.e. the ability to detect unstable conditions, as in previous studies, but a much lower probability of null events (0.54–0.75, PON), i.e. the ability to detect stable conditions, with variation due to the binary classification of "Fair" stability as stable or unstable. Adding a 2 m test after an ECTP result lowered the POD (0.49–0.58), but substantially raised the PON (0.88–0.98) of the combined tests. The proportion of tests in agreement, i.e. ECTP and 2 m ECTP, increases with decreasing stability. We conclude that an ECTP followed by a 2 m ECTP is a clear red flag, indicating instability. The interpretation of an ECTP followed by a 2 m ECTN/X (no propagation) is not clear. Though this result suggests stability, a high potential for a false stable result means we cannot recommend the 2 m ECT for binary stability assessments.

1. Introduction

Stability tests are one of the most commonly used methods to assess avalanche likelihood. The concept is to attempt to simulate a small failure that can be correlated to slope scale avalanche danger. Stability tests involve loading an isolated column or beam of snow, primarily to intensify stress, such that the test will fail before the slope, given similar loads.

The Extended Column Test (ECT, Simenhois and Birkeland, 2009) has become the most widely used stability test among SnowPilot users (Birkeland and Chabot, 2012), who are mostly in the US. Briefly, the Extended Column Test involves a 30 cm upslope by 90 cm cross slope beam that is isolated from the surrounding snowpack and then loaded by placing a shovel on top of the beam and tapping from the wrist, elbow, and then shoulder.

In a comparison of stability tests (Schweizer and Jamieson, 2010), the Extended Column Test (ECT) had the highest unweighted average accuracy (UAA), an average of the probability of detection (POD) and the probability of a null event (PON, Doswell et al., 1990). The UAA for the ECT was 89% over all data sources. In comparison, UAAs were 79% for the Propagation Saw Test (PST, Gauthier and Jamieson, 2008a) and 68% for the Compression Test (CT, with fracture character; van Herwijnen and Jamieson, 2007). Like all tests, the ECT is not perfect and there may be ways to improve it. Not all studies have found the ECT to be such an accurate stability test. Ross (2010) reports an UAA of only 58% for the ECT. Hendrikx et al. (2009) report that 34–50% of ECTs propagated in grids done on the same slopes, suggesting an UAA around 50%.

When ECT guidelines were developed, its 90 cm length was not extensively tested against other beam lengths. Recently, Bair et al. (2014) showed that in PSTs, shorter beams have higher energy release rates for common critical crack lengths r_c, 20–40 cm (Gauthier and Jamieson, 2008a; Ross and Jamieson, 2012). They concluded the higher rates were caused by stress intensification from the far edge of the beam and that ECTs were subject to the same edge effect, given that the initial cracked area in an ECT is similar to common critical crack lengths (van Herwijnen and Birkeland, 2014). By far edge, we refer to the edge furthest from the section of the beam being loaded (Figure 2 in Bair et al., 2014). For beams ≥ 2 m, the energy release rates become asymptotic for common r_c values. This finding was unaffected by linear elastic or viscoelastic assumptions. Thus, we suggest that shorter tests could reach the critical energy release rate when longer tests may not reach

* Corresponding author at: 6832 Ellison Hall University of California Santa Barbara, CA 93106-3060, USA.
E-mail address: nbair@eri.ucsb.edu (E.H. Bair).

http://dx.doi.org/10.1016/j.coldregions.2015.06.021

the critical rate. This edge effect could lead to propagation (ECTP) in shorter tests, but not in longer tests for identical snowpacks.

Another reason why a 2 m test might propagate fully less often compared to a standard length test is the tensile strength of the slab. In 2 m tests, slab fracture may occur more often and arrest the crack in the weak layer. In standard length tests, slab fracture may not occur as often since there is less tensile stress in the slab. The increased tensile stress arises from longer unsupported sections of the beam in the 2 m tests during crack propagation. Gaume et al. (2015) used a discrete element model of the PST to show that a slab strength of 6–16 kPa, or a slab density of 200–300 kg m^{-3} using the model of Sigrist (2006), is required to prevent slab fracture. Slabs with strengths below this critical value may fracture and arrest the weak layer crack in 2 m tests, but not in standard length tests. Again, we suggest this finding is applicable to the ECT as well.

2. Methods

Given the finding that 2 m ECTs eliminate the far edge effect on propagation (Bair et al., 2014), we tested the added value of 2 m tests performed after standard length tests that propagated. Tests were performed in 2013–2014 by eight avalanche professionals in the US (California, Alaska, Nevada, and Montana) and by one professional in Switzerland. Observers were given instructions to record an a priori slope stability rating. Information for this rating could come from any source except a stability test done on that slope on that day. This provision ensured that observers would not base their stability rating on test results. The a priori stability rating is based on a five point scale: “Very Poor”, “Poor”, “Fair”, “Good”, and “Very Good” (Greene et al., 2010). These five choices offer a more detailed assessment of slope stability than the binary “stable/unstable” rating that has been used in previous studies (Gauthier and Jamieson, 2008b; Schweizer and Jamieson, 2010; Simenhois and Birkeland, 2009; Winkler and Schweizer, 2009).

After recording an a priori stability rating, observers performed a standard length 90 cm ECT (Simenhois and Birkeland, 2009). If the standard length test propagated (ECTP), a second 2 m long test was performed, adjacent to the first test. Because of the decreased stress at the crack tip in the longer test, we assumed that if the standard length test did not propagate (ECTN/X), the 2 m ECT would not propagate either. Thus, no further tests were done after an ECTN/X. This assumption and its consequences are covered in detail in Section 4.

To evaluate our tests, we used 2 × 2 contingency tables (Doswell et al., 1990) based on stable and unstable classifications for tests and for stability ratings. These classifications oversimplify the problem, but they are widely used and provide a means for comparison with other studies. Given that “Fair” could be classified as stable or unstable, we made comparisons where: 1) a slope was classified as unstable if its stability rating was “Very Poor” or “Poor” and stable otherwise; or 2) a slope was classified as unstable if its stability rating was “Very Poor”, “Poor”, or “Fair” and stable otherwise. An ECT was classified as unstable if it was an ECTP and stable otherwise. When the 2 m ECT was included, an ECTP followed by a 2 m ECTP was classified as unstable. An ECTN/X or an ECTP followed by a 2 m ECTN/X was classified as stable.

Table 1
Contingency table for ECTs, with “Fair” stability indicating a stable slope. A stable test result is an ECTN/X. An unstable test result is an ECTP. A stable observation is “Fair”, “Good”, or “Very Good” stability. An unstable observation is “Poor” or “Very Poor” stability.

		Observations	
		Stable	Unstable
Tests	Stable	78	6
	Unstable	67	69

Table 2
Contingency table for ECTs, with “Fair” stability indicating an unstable slope. Stable and unstable test results are the same as in Table 1 with the difference being that a stable observation here is “Good” or “Very Good” stability and an unstable observation is “Fair”, “Poor”, “Very Poor” stability.

		Observations	
		Stable	Unstable
Tests	Stable	68	16
	Unstable	23	113

Contingency table statistics (Doswell et al., 1990) computed were: the probability of detection (POD),

$$\mathrm{POD} = \frac{\text{correct unstable prediction}}{\text{all observed unstable slopes}}, \tag{1}$$

the probability of a null event (PON),

$$\mathrm{PON} = \frac{\text{correct stable prediction}}{\text{all observed stable slopes}}, \tag{2}$$

the false alarm ratio,

$$\mathrm{FAR} = \frac{\text{incorrect unstable prediction}}{\text{all predicted unstable slopes}}, \tag{3}$$

and the average of the POD and the PON, the unweighted average accuracy (UAA, Schweizer and Jamieson, 2010)

$$\mathrm{UAA} = \frac{\mathrm{PON} + \mathrm{POD}}{2}. \tag{4}$$

Tests with a high POD correctly identify unstable slopes most of the time, while tests with a high PON correctly identify stable slopes most of the time. An ideal test will have a high PON and POD, but most tests are unbalanced. For snow stability tests, unbalanced tests with POD > PON are preferable to those with PON > POD since the consequences of a false stable prediction are more severe than those of a false unstable prediction (Schweizer and Jamieson, 2010).

To test for differences between the ECTP and the 2 m ECT groups, statistical tests were used. These tests were: the nonparametric Kolmogorov–Smirnov test (KS test, Massey, 1951), the Anderson–Darling test (AD test, Anderson and Darling, 1952) for normality, and the Student's *t* test (Box, 1987).

3. Results

We used results from 220 ECTs, including 136 ECTPs followed by 2 m ECTs. The contingency tables for ECTs (Tables 1 and 2) can be compared to the contingency tables for ECTP/2 m ECTs (Tables 3 and 4) and a table of statistics for both (Table 5) summarizes the findings. The number of tests and the number of unstable slopes (*N* and *base rate* in Table 5) are the same for the ECT and the ECTP/2 m ECT and these values are comparable to previous studies (summarized in Schweizer and Jamieson, 2010).

Table 3
Contingency table for ECTs followed by 2 m ECTs, with “Fair” indicating a stable slope. Stable and unstable observations are the same as in Table 1. A stable test result pair is an ECTN/X or an ECTP followed by a 2 m ECTN/X. An unstable test result pair is an ECTP followed by a 2 m ECTP.

		Observations	
		Stable	Unstable
Tests	Stable	127	37
	Unstable	18	38

Table 4
Contingency table for ECTs and ECTPs followed by 2 m ECTs, with “Fair” indicating an unstable slope. Stable and unstable observations are the same as in Table 2. Stable and unstable test pairs are the same as in Table 3.

		Observations	
		Stable	Unstable
Tests	Stable	89	75
	Unstable	2	54

Of particular note is the high POD for the ECT (0.88–0.92) and the low PON (0.54–0.75), with the variation due to the binary classification of “Fair” stability. These results show that the ECT was quite accurate at detecting unstable conditions, but much less accurate at detecting stable conditions. The ECT also had a high FAR (0.17–0.49), the prediction of an unstable slope when it was observed stable. In contrast, adding a 2 m ECT yielded a low POD (0.49–0.58), but a high PON (0.88–0.98) and a lower FAR (0.04–0.32). We interpret this imbalance to indicate that adding a 2 m ECT after an ECTP was not accurate for detecting unstable conditions, but very accurate at detecting stable conditions. The imbalances nearly offset one another, resulting in comparable UAA values, 0.73–0.81 for the ECT and 0.68–0.78 for the ECTP/2 m ECT.

The ECT alone shows a clear trend of increasing propagation likelihood with decreasing stability (Fig. 1), with 9% propagation (ECTP) in “Very Good” stability to 100% propagation in “Very Poor” stability. There is a large increase in propagation, 31–81%, from “Good” to “Fair” stability. All of the stability classes had more than 20 observations.

For the ECTP/2 m ECT, there is a trend of increasing agreement as stability decreases (Fig. 2). Tests in agreement (ECTP and 2 m ECTP) increase from: 0% for “Very Good”, to 10% for “Good”, to 36% for “Fair”, to 56% for “Poor”, to 55% for “Very Poor”. All of the stability classes had more than 20 observations except “Very Good” which only had 2 pairs of ECTP and 2 m ECTs. Of note for the ECT alone and the paired tests is that 41/42 ECTs/paired tests in “Very Poor” came from tests done at avalanches, with 38 pairs done between 10 min and 2 h after the avalanche.

The 2 m ECTs took significantly more taps to fail (t test $p < 0.001$), with the difference in taps being normally distributed (AD test $p < 0.001$). The median number of taps for an ECTP was 17, but the median for a 2 m ECT (with or without propagation) on the same layer as the ECTP was 20 taps (Fig. 3). Propagation in the 2 m ECTs depended on slab thickness. The median slab thickness (slope normal) was 43 cm for 2 m ECTP but only 27 cm for 2 m ECTN/X (Fig. 4). The two groups were statistically different (KS test $p < 0.01$). The median propagation distance for ECTP followed by a 2 m ECTN was 108 cm, measured from the trigger edge (Fig. 5). Thus, cracks traveled slightly further than the length of a standard ECT, 90 cm. Other snow profile variables (e.g. slab/weak layer hardness, crystal type, crystal size) did not show a relationship to propagation in the 2 m ECT. We did not have enough slab density measurements for comparison.

Table 5
Test statistics from Tables 1–4. N is the total number of tests. The base rate is the proportion of unstable slopes. The POD is the probability of detection. The PON is the probability of a null event. The FAR is the false alarm ratio. The UAA is the unweighted averaged accuracy. For definitions see Section 2.

	With “Fair” as stable		With “Fair” as unstable	
	ECT	ECTP followed by 2 m ECT	ECT	ECTP followed by 2 m ECT
N	220	220	220	220
Base rate	0.34	0.34	0.59	0.59
POD	0.92	0.49	0.88	0.58
PON	0.54	0.88	0.75	0.98
FAR	0.49	0.32	0.17	0.04
UAA	0.73	0.68	0.81	0.78

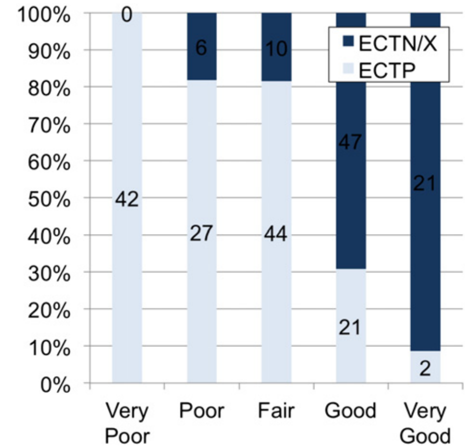

Fig. 1. Percentage of ECTs that propagated vs. those that did not, grouped by a priori stability rating. The black numbers in each bar are the numbers of tests in each category.

4. Discussion

As mentioned in Section 2, we assumed a longer test would not propagate if a shorter test did not because of decreased stress at the crack tip in the longer test. We tested and confirmed this assumption informally with about ten 2 m tests following ECTN/X results throughout the winter. We never observed propagation in the 2 m test after an ECTN/X. Since a 2 m test requires about twice as much excavation as a standard length test, we decided that these guidelines were the most efficient for observers with limited time to perform stability tests. The greater number of taps (Fig. 3) and the greater slab thickness (Fig. 4) needed for failure in the 2 m ECT compared to the ECT further support this assumption by showing that the 2 m ECT is more difficult to trigger. The ECT has high repeatability (87% in Winkler and Schweizer, 2009) and given the lower propagation likelihood of the 2 m ECT compared to the ECT (Section 3), we assume the likelihood of a 2 m ECTN/X following an ECTN/X would be at least as great and probably greater.

Yet, we note that our assumption is untested except for the informal tests and point out to the reader that this is a limitation in our study. The result of this assumption is that this study was not able to directly compare the performance of the ECT and the 2 m ECT. Instead, we assessed the added benefit of performing a 2 m ECT after an ECTP.

Another option to reduce edge effects would have been to perform stability tests with far edges that were not isolated instead of longer tests. We decided against this option. In our experience and in others'

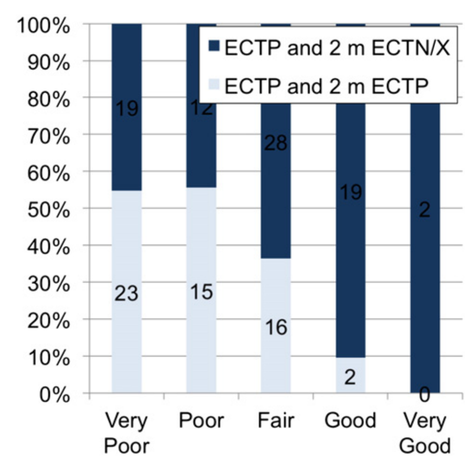

Fig. 2. Percentage of standard length and 2 m ECTs in agreement, grouped by a priori stability rating. The black numbers in each bar are the numbers of paired tests in each category.

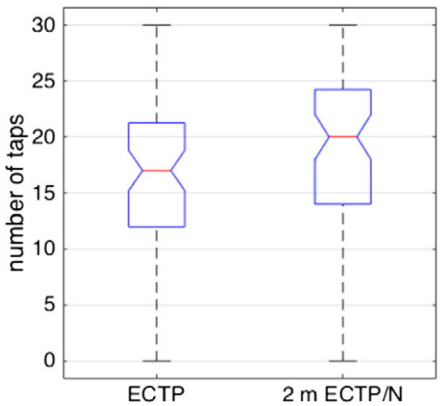

Fig. 3. The number of taps for a 2 m ECTP/N that failed on the same layer as an ECTP. Median is red, the 25th–75th percentile is blue, and the whiskers are ranges. The notched area signifies a 95% confidence interval for the median. Color figures are provided in the online version of this article.

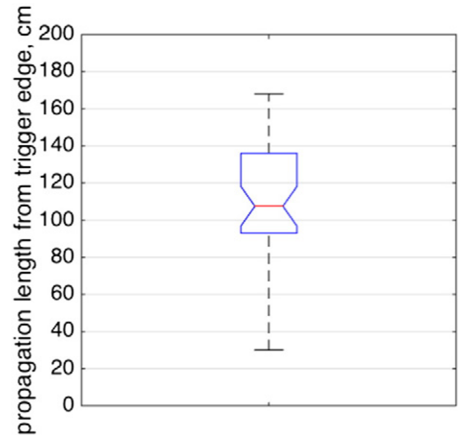

Fig. 5. Crack propagation length for 2 m ECTN, following ECTP. Measured from the trigger edge of the beam, N = 40. Box plot features are the same as in Fig. 3.

(Ross and Jamieson, 2012), it can be difficult to discern how far a crack travels in such tests, a critical piece of our research. For instance, cracks in the weak layer can disappear into the uncut end without causing a fracture through the slab, leaving the tester wondering how much further past the end of the beam the crack traveled.

The POD for the ECT (0.88–0.92), is comparable to most previous studies (0.83–1.00, Moner et al., 2008; Simenhois and Birkeland, 2009; Winkler and Schweizer, 2009). However, the PON (0.54–0.75) was much lower compared to those same studies (0.79–0.97). These results beg the question: why is the PON for the ECT substantially lower in this study than in most previous studies?

To address this question, we examined potential observer bias. The proportion of false unstable ECT results relative to the total number of unstable ECT results varied greatly with observer and location, from 0–90% (Figs. 6 and 7), again dependent on the binary classification of "Fair" stability. There is a trend of a high number of false unstable observations from observers (a), (b), (c), and (d) in the Sierra Nevada (CA and CA/NV in Fig. 6), accounting for 57/75 (76%) of the total false stable results when "Fair" is classified as unstable. This trend disappears when "Fair" is classified as "stable" (Fig. 7), showing that the binary classification of "Fair" stability largely impacts whether a test is a true or false unstable. Nonetheless, for either classification of "Fair" stability, tests from CA and NV still had the highest number and percentage of false unstable results. We suggest that the Sierra Nevada had snowpack conditions in 2013–14 that caused high false unstable rates. A severe drought caused well below average snow depths and a thick layer of basal depth hoar that showed ECT propagation for much of the season. Yet, stability was often "Fair" or better. Thus, this thick layer of depth hoar and a shallow snowpack may have been especially conducive to producing false unstable results in the ECT. Given that basal depth hoar is not unusual, especially in the western US, we suggest that these false unstable observations from California are still valid for evaluating the ECT and the 2 m ECT.

One study (Ross, 2010) found a lower POD and PON (0.57 and 0.37, respectively) for the ECT than this study. The author of that study speculates that a wider range of slab depths and softer snow tested in his study compared to other studies (particularly Simenhois and Birkeland, 2009) may be the cause of the relatively low rates. It is possible that our study performed ECTs in snowpack conditions that may not have been thoroughly tested in previous studies.

Our results suggest that the ECT alone is preferable to an ECTP followed by a 2 m ECT for binary stability assessments. As discussed in Section 2, imbalanced stability tests with POD > PON are preferable to those with PON > POD because the former promote cautious decisions. On that basis, the ECT is preferred since it has a much lower false stable rate (1-POD), 0.08–0.12. However, binary stability assessments oversimplify the avalanche problem. For more complex stability assessments, we suggest that a 2 m ECT after an ECTP provides added stability information, especially if the user suspects a false unstable

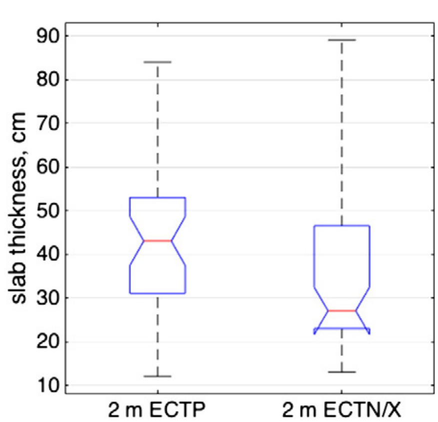

Fig. 4. Slab thickness for 2 m tests with (ECTP) and without (ECTN/X) propagation; N = 38 ECTP and 47 ECTN/X where slab thickness (slope normal) was measured. Box plot features are the same as in Fig. 3.

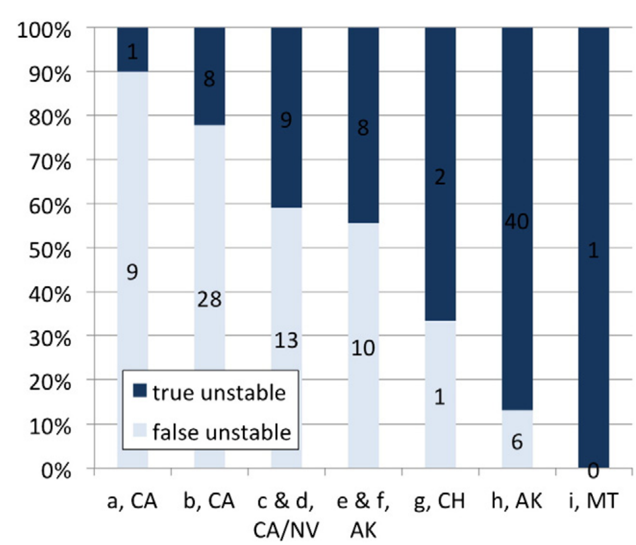

Fig. 6. Proportion of unstable ECT results sorted by observer, with "Fair" as stable. True unstable results are ECTP with "Very Poor" or "Poor" stability. False unstable results are ECTP with "Fair", "Good", or "Very Good" stability. The numbers in each bar are the counts of true/false unstable results. Observers are labeled by letters with their location indicated by their two letter US state abbreviation. Switzerland is CH.

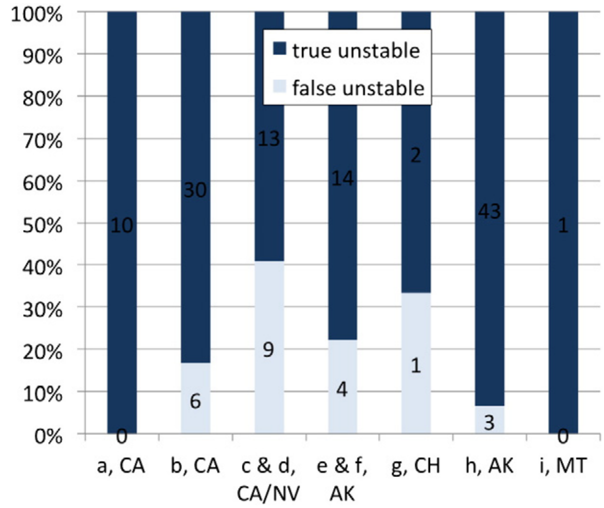

Fig. 7. Proportion of unstable ECT results sorted by observer, with "Fair" as unstable. True unstable results are ECTP with "Very Poor", "Poor", or "Fair" stability. False unstable results are ECTP with "Good" or "Very Good" stability. The rest is the same as in Fig. 6.

result from the first test, though this should be done with caution. A user may suspect a false unstable result because of information from other sources, such as a lack of avalanche activity or lack of other signs of instability.

We suggest that the increasing agreement (ECTP and 2 m ECTP) as stability decreases suggests that propagation in both tests is a clear red flag or a strong sign of instability. The interpretation of an ECTP followed by a 2 m ECTN/X is not clear. Even in "Poor" and "Very Poor" stability, 44 and 45% of 2 m ECTs did not propagate. This is reflected in the high false stable rate (1-POD or 0.42–0.51) of the 2 m ECT.

The low POD and the high proportion of stable test results in the "Very Poor" and "Poor" categories are a problem for the 2 m ECT. Although we note that much of the verification was done at avalanches. In the "Very Poor" category, which represented 33–56% of the unstable observations, 41/42 (98%) of the pairs came from avalanche sites. Avalanche sites have disturbed snowpacks that may extend beyond the perimeter of the slab. Additionally, the most unstable snow on the slope has already slid and the stability can change rapidly, especially with weak layers of precipitation particles. Thus, there are problems with verifying stability tests at avalanche sites. For example, at avalanche accident sites the POD for the ECT is 0.64–0.75 (Fig. 8), below the range reported here and in the majority of previous studies (0.83–1.00, Moner et al., 2008; Simenhois and Birkeland, 2009; Winkler and Schweizer, 2009).

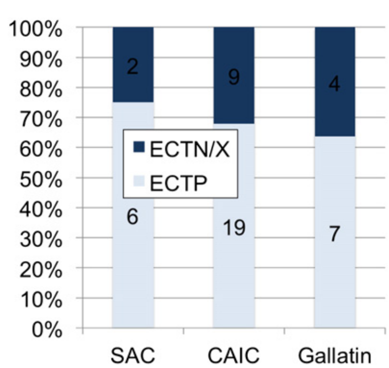

Fig. 8. ECTs at avalanche accidents. Data from the Sierra Avalanche Center (SAC), the Colorado Avalanche Information Center (CAIC), and the Gallatin National Forest Avalanche Center (Gallatin).

5. Conclusion

Previous work (Bair et al., 2014) suggests that the standard 90 cm Extended Column Test (ECT) may propagate when longer tests will not because of a far edge effect. Models from this work show that this far edge effect disappears for tests ≥2.0 m in length. To test the accuracy of a 2 m ECT, we collected data on 220 ECTs, 136 of which were done side-by-side with 2 m ECTs, using an a priori stability rating for verification. We assumed that an ECTN/X (no propagation, stable) would be followed by a 2 m ECTN/X; thus we only performed 2 m ECTs after an ECTP (propagation, unstable). This assumption was supported by our findings. The 2 m ECT required deeper slabs and more taps to fail than the ECT, indicating a test that required greater load to fail.

Our results showed a similar probability of detection (POD, i.e. the ability to detect unstable conditions) for the ECT as in previous studies, but a much lower probability of a null event (PON, i.e. the ability to detect stable conditions). For the 2 m ECT following an ECTP, the results were reversed with a low POD and a high PON. We interpret these results to show that the ECT was effective at detecting unstable conditions, but had a high likelihood of false unstable results. Including a 2 m ECT after an ECTP was effective at detecting stable conditions, but had a high likelihood of false stable results. Additionally, the proportion of tests in agreement (ECTP and 2 m ECTP) increased with decreasing stability. Still, even at "Poor" and "Very Poor" stability, 44 and 45% of the 2 m ECTs did not propagate. Given that the potential consequences of a false stable result are more severe than those of a false unstable result, we conclude that the ECT is superior to the 2 m ECT for binary stability assessments. Therefore, we do not recommend the 2 m ECT for this purpose.

Acknowledgments

We thank two anonymous reviewers for their helpful comments. We thank Brandon Schwartz and Andy Anderson at the Sierra Avalanche Center, Sue Burak at the Eastern Sierra Avalanche Center, and Mike Janes at Alaska Electric Light and Power for their stability tests. The first author was partly supported by a US Army Corps of Engineers Cold Regions Research and Engineering fellowship administered by ORISE/ORAU.

References

Anderson, T.W., Darling, D.A., 1952. Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. pp. 193–212. http://dx.doi.org/10.1214/aoms/1177729437.

Bair, E.H., Simenhois, R., van Herwijnen, A., Birkeland, K., 2014. The influence of edge effects on crack propagation in snow stability tests. Cryosphere 8 (4), 1407–1418. http://dx.doi.org/10.5194/tc-8-1407-2014.

Birkeland, K.W., Chabot, D., 2012. Changes in stability test usage by Snowpilot users. Proceedings of the 2012 International Snow Science Workshop, Anchorage, AK USA, pp. 1065–1068.

Box, J.F., 1987. Guinness, Gosset, Fisher, and small samples. Stat. Sci. 2 (1), 45–52. http://dx.doi.org/10.2307/2245613.

Doswell, C.A., Davies-Jones, R., Keller, D.L., 1990. On summary measures of skill in rare event forecasting based on contingency tables. Weather Forecast. 5 (4), 576–585. http://dx.doi.org/10.1175/1520-0434(1990)005<0576:OSMOSI>2.0.CO;2.

Gaume, J., van Herwijnen, A., Chambon, G., Schweizer, J., Birkeland, K.W., 2015. Modeling of crack propagation in weak snowpack layers using the discrete element method. Cryosphere Discuss. 9 (1), 609–653. http://dx.doi.org/10.5194/tcd-9-609-2015.

Gauthier, D., Jamieson, B., 2008a. Evaluation of a prototype field test for fracture and failure propagation propensity in weak snowpack layers. Cold Reg. Sci. Technol. 51 (2–3), 87–97. http://dx.doi.org/10.1016/j.coldregions.2007.04.005.

Gauthier, D., Jamieson, B., 2008b. Fracture propagation propensity in relation to snow slab avalanche release: validating the Propagation Saw Test. Geophys. Res. Lett. 35 (13), L13501. http://dx.doi.org/10.1029/2008gl034245.

Greene, E., Atkins, D., Birkeland, K.W., Elder, K., Landry, C., Lazar, B., McCammon, I., Moore, M., Sharaf, D., Sterbenz, C., Tremper, B., Williams, K., 2010. Snow, Weather, and Avalanches: Observational Guidelines for Avalanche Programs in the United States. p. 136.

Hendrikx, J., Birkeland, K., Clark, M., 2009. Assessing changes in the spatial variability of the snowpack fracture propagation propensity over time. Cold Reg. Sci. Technol. 56 (2–3), 152–160. http://dx.doi.org/10.1016/j.coldregions.2008.12.001.

Massey, F.J., 1951. The Kolmogorov–Smirnov test for goodness of fit. J. Am. Stat. Assoc. 46 (253), 68–78. http://dx.doi.org/10.1080/01621459.1951.10500769.

Moner, I., Gavaldà, J., Bacardit, M., Garcia, C., Martí, G., 2008. Application of field stability evaluation methods to the snow conditions of the Eastern Pyrenees. International Snow Science Workshop, Whistler, BC Canada, pp. 386–392.

Ross, C., 2010. Testing Fracture Propagation Propensity for Slab Avalanche Forecasting. University of Calgary, Calgary, AB Canada (199 pp.).

Ross, C.K.H., Jamieson, B., 2012. The propagation saw test: slope scale validation and alternative test methods. J. Glaciol. 58 (208), 407–416. http://dx.doi.org/10.3189/2012JoG11J192.

Schweizer, J., Jamieson, B., 2010. Snowpack tests for assessing snow-slope instability. Ann. Glaciol. 51 (54), 187–194.

Sigrist, C., 2006. Measurement of Fracture Mechanical Properties of Snow and Application to Dry Snow Slab Avalanche Release (Diss. No. 16736). http://dx.doi.org/10.3929/ethz-a-005282374.

Simenhois, R., Birkeland, K., 2009. The extended column test: test effectiveness, spatial variability, and comparison with the propagation saw test. Cold Reg. Sci. Technol. 59 (2–3), 210–216. http://dx.doi.org/10.1016/j.coldregions.2009.04.001.

van Herwijnen, A., Birkeland, K.W., 2014. Using high-speed video to better understand extended column tests. Cold Reg. Sci. Technol. 97, 97–103. http://dx.doi.org/10.1016/j.coldregions.2013.07.002.

van Herwijnen, A., Jamieson, B., 2007. Fracture character in compression tests. Cold Reg. Sci. Technol. 47 (1–2), 60–68. http://dx.doi.org/10.1016/j.coldregions.2006.08.016.

Winkler, K., Schweizer, J., 2009. Comparison of snow stability tests: Extended Column Test, Rutschblock Test and Compression Test. Cold Reg. Sci. Technol. 59 (2), 217–226. http://dx.doi.org/10.1016/j.coldregions.2009.05.003.