COMPARISON OF DIFFERENT STABILITY TESTS REGARDING THEIR UTILITY FOR SNOWPACK STABILITY ASSESSMENTS IN GUIDING OPERATIONS

Anna Heuberger

University of Natural Resources and Life Sciences, Vienna

2021

COMPARISON OF DIFFERENT STABILITY TESTS REGARDING THEIR UTILITY FOR SNOWPACK STABILITY ASSESSMENTS IN GUIDING OPERATIONS

by
Anna Heuberger

B.Sc., University of Innsbruck, 2016
B.A., University of Innsbruck, 2015

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

Diplom-Ingenieurin

in

Alpine Naturgefahren

UNIVERSITY OF NATURAL RESOURCES AND LIFE SCIENCES, VIENNA
2021

Affidative

I hereby declare that I am the sole author of this work. No assistance other than that which is permitted has been used. Ideas and quotes taken directly or indirectly from other sources are identified as such.

This work has not been submitted in the same or similar form to any other examiners as a form of examination.

Saalfelden, April 7th, 2021

Anna Heuberger

Abstract

Dry slab avalanches are the main safety concern for commercial skiing operations in Canada. Weather data and public avalanche bulletins are often limited due to the remote and vast nature of the operating tenures. Therefore, operators largely depend on snowpack stability tests to assess snowpack stability and avalanche hazard. With limited temporal resources during operations, guides often use small-scale stability tests (e.g. Shovel Test (ST) and Compression Test (CT)) to quickly extract valuable information from the snowpack and by performing a high number of such tests every day, receive a bigger picture of the snowpack in the area. However, there is a current lack of knowledge on the value of small-scale tests as a tool to assess stability and whether differences in the testing methods depend on snowpack properties such as the weak layer grain type or size. This study uses two datasets collected during the 2019/2020 winter season to perform a statistical analysis aimed at answering the questions in place. The first dataset consists of 633 test comparisons including CT and ST results collected by guides. The objectives of this dataset are to evaluate if there is a correlation between the test results and the stability rating and to determine whether differences in the two testing methods depend on certain snowpack properties. The analysis shows that both tests correlate significantly with the stability rating and that the comparison results are dependent on snowpack properties such as the depth of the weak layer.

The second dataset is aimed at determining the accuracy of the fracture character of small-scale tests as a tool to predict propagation propensity as measured in the Extended Column Test (ECT) and the Propagation Saw Test (PST). This Dataset is comprised of 138 test comparisons including the CT, ST, ECT and PST. The data shows that the CT is likely to produce sudden results on layers without propagation propensity while the ST fails to produce Sudden results on critical layers. This difference between the results of the CT and the ST is most prominent for storm snow instabilities and weak layer/slab interfaces with a large hardness difference.

Keywords: Stability Test, snowpack stability, avalanche hazard, snowpack properties, fracture character, compression test, shovel shear test, mechanized skiing, helicopter skiing.

Kurzfassung

Trockene Schneebrettlawinen stellen das größte Risiko für Gäste und Skiführer/innen kommerzieller Helikopter-Ski- Unternehmen in Kanada dar. Wetterdaten und öffentliche Lawinenlageberichte haben wegen der Weitläufigkeit und der Abgeschiedenheit der Gebiete nur einen eingeschränkten Wert für die lokale Gefahrenbeurteilung. Auf Grund der begrenzten zeitlichen Ressourcen während des Betriebes, verwenden Skiführer/innen oft kleinflächige Stabilitätstests wie den Kompressionstest (CT) oder den Schaufelschertest (ST) um effizient wertvolle Informationen aus der Schneedecke zu gewinnen und so durch eine hohe Anzahl kleiner Tests einen Überblick über die Gesamtsituation zu erlangen. Nichtsdestotrotz, ist noch wenig darüber bekannt, wie geeignet kleinflächige Stabilitätstests für die Stabilitätsvorhersage sind und wie unterschiedliche Ergebnisse von Schneedeckencharakteristika wie der Korngröße oder der Kornform der Schwachschicht abhängen.

Um diese Fragen zu beantworten, wurden während der Wintersaison 2019/2020 Daten erhoben und statistisch analysiert. Der erste Datensatz besteht aus 633 Testvergleichen zwischen dem CT und dem ST welche von Skiführern durchgeführt wurden. Die Analyse dieses Datensatzes zeigt, dass die Ergebnisse der zwei Tests signifikant mit der Stabilitätseinschätzung korrelieren und dass die Unterschiede zwischen den Testmethoden von Schneedeckeneigenschaften, wie der Tiefe der Schwachschicht, abhängen.

In der Analyse des zweiten Datensatzes wird untersucht, inwiefern die Bruchcharakteristik der kleinflächigen Stabilitätstests das Fortpflanzungspotential des Bruchs vorhersagen kann. Das Fortpflanzungspotenzial wird dabei durch den erweiterter Säulentest (ECT) und durch den *Propagation Saw Test* (PST) bestimmt. Dieser Datensatz besteht aus 138 Testvergleichen bei denen der CT, ST, PST und ECT inkludiert sind. Aus diesen Datensatz geht hervor, dass der CT mit hoher Wahrscheinlichkeit regelmäßige Bruchflächen (*Sudden*) auf Schwachschichten ohne Fortpflanzungspotenzial produziert während der ST auf wenig ausgeprägten Schwachschichten mit Fortpflanzungspotenzial keine regelmäßigen Bruchflächen erzeugt. Diese Unterschiede kommen besonders zur Geltung, wenn es sich um Neu- oder Triebschneeprobleme handelt oder wenn ein großer Härteunterschied zwischen der Schwachschicht und dem darüber liegenden Brett besteht.

Acknowledgements

This piece of work would not have been possible without the help of my supervisor Dr. Ingrid Reiweger and the support from Mike Wiegele Heli-Skiing.
Thank you, Ingrid, for sharing your advice and expertise with me. It was a real pleasure to learn from your experience, curiosity and knowledge over the past years.
The idea for this thesis was born while discussing research topics with Mike Wiegele. Fired by his relentless passion and curiosity for snow and avalanches and his stubborn determination to dig deeper. Thank you for all your support, the opportunities and inspiration Mike.
The realization of the data collection with the logistics and organization involved would not have been possible without Elias Ortner. Thank you, Elias, for puzzling the pieces together. Your drive to get things done is contagious which thankfully helped me push through the data analysis.
Of course, I would also like to thank the great team of guides at MWHS for your advice, patience and support with this project. Not to forget the pilots who gave us lifts to and from the research areas alongside long days of work.

Matthias, who supported me with this research from the beginning. Thank you for always putting up with me, for your motivation and for offering your advice and expertise when needed.

Last but certainly not least my parents. For always supporting me.

Table of Contents

List of Tables

List of Figures

List of Acronyms

CAA	Canadian Avalanche Association
CT	Compression Test
ECT	Extended Column Test
DF	Decomposing Fragments
FC	Facets
FCxr	Facets on crust combination
MWHS	Mike Wiegele Heli-Skiing
OGRS	Observation Guidelines and Recording Standards for Weather, Snowpack and Avalanches
PP	Precipitation Particles
PPgp	Graupel
PST	Propagation Saw Test
RB	Rutschblock
SH	Surface Hoar
ST	Shovel Test

1 Research background and problem setting

The biggest risk in helicopter skiing in Canada is dry slab avalanches (Figure 1). Even though the risk has decreased substantially over the last decades, it still accounts for 85% of the overall risk of death involved (Walcher, Haegeli, and Fuchs 2019). Therefore, this is the prominent safety concern in guiding operations involved in commercial helicopter backcountry skiing.

Figure 1: Naturally released dry slab avalanche on surface hoar.

The prerequisites for a dry slab avalanche are a snowpack which contains a weak layer with a cohesive layer of snow on top of it and sufficient slope inclination. For the avalanche to release a fracture needs to initiate and then propagate through the weak layer. In the case of a skier triggered avalanche the fracture initiates by the formation of a crack in the weak layer as a result from the skier impact. If the initiated crack reaches a critical size, which is linked to properties of the slab overlying the weak layer, rapid crack propagation starts (Schweizer, Mccammon, and Jamieson, 2008). Snowpack stability is therefore determined by the shear and compression strength of the weak layer and the properties of the planar slab on top of it (McClung and Schweizer, 2006).

To determine snowpack stability and further draw conclusions on avalanche hazard, ski-guides perform so-called stability tests to support their assessment. Thereby, an isolated column of the snowpack is stressed to determine the strength of existing weak layers (failure initiation) and to draw conclusion on its propagation potential.

In the past decades, a variety of different testing methods have been established with the aim to support the stability assessment on a slope (e.g. Jamieson and Ross, 2012). Comparisons of such tests revealed, that stability tests with a bigger surface area, such as the Rutschblock test (Föhn, 1986) or the Extended Column Test (Birkeland and Simenhois, 2006) tend to have a higher accuracy in forecasting the actual stability of a slope than smaller testing methods (Winkler and Schweizer, 2008).
Especially for guides travelling through avalanche terrain with guests however, it is crucial to extract valuable information from testing sites in a limited amount of time. Therefore, time-efficient tests like the Compression Test (CT; Jamieson and Johnston, 1996) and the Shovel Shear Test (ST; Schaerer, 1988) are frequently used to obtain information from various point observations and receive a bigger picture of snowpack stability.
Compared to tests with larger test areas (e.g. RB, ECT), smaller tests have the disadvantage that they do not allow direct observation of the propagation propensity of a weak layer after failure is initiated. The propagation potential is only indirectly assessed based on the fracture character. Furthermore, in smaller scale tests weak layers are more sensitive to stress which leads to more results in less relevant weak layers (Schweizer and Jamieson, 2010). In existing literature the ST is described as subjective and hard to interpret (Tremper, 1994) and it is recommended to repeat the test at the same study location to draw reliable conclusions on a weak layer's stability (McClung and Schaerer, 2006).

Existing research has compared different stability tests and related test results to snowpack characteristics (e.g. Ross and Jamieson, 2008). However, no study has completed such research including all four stability tests, the CT, ST, ECT and PST, to evaluate which testing method provides the most accurate results for *known* weak layers under which circumstances (snowpack/slab properties). Since CTs and STs are still common practice in operational heliskiing in Canada, this study investigates which test method gives the best results for different weak layer and slab characteristics and which deficiencies the testing methods have.

In a second part of the study, a test series including ECTs, Propagation Saw Tests (PST; (Gauthier and Jamieson, 2008) as well as STs and CTs will be conducted. Past research (e.g. Johnson and Birkeland, 2002, Herwijnen and Jamieson, 2004) proposes, that *Sudden (Sudden* *collapse or Sudden Planar)* fracture characters are an indication that slab avalanches are likely to propagate if a crack occurs in the weak layer . The authors of this study used skier testing or/and field and snowpack observations indicating instabilities (e.g. whumpfs, cracking, remote triggering) to determine propagation propensity of the tested layers. The study at hand explores how the fracture character in CTs and STs correlates with the propagation potential of the weak layers as indicated by the ECT or PST and how this in return depends on the weak layer properties.

Objectives and hypothesis

This study compares test results done in an operational guiding setting (CT, ST) as well as in a research setting (CT, ST, ECT, PST) where the aspect of time played an inferior role in the completion of the tests.

The aim is to (1) analyze the suitability and validity of small-scale tests for quickly assessing snowpack stability in an operational guiding setting. Therefore, a large number of CT and ST results performed by guides during operations are compared regarding properties of the snowpack layering and the stability rating in order to compare the accuracy and meaningfulness of the results. This allows us to draw conclusions under which circumstances these tests might deliver the most meaningful information for the user.

The second aim (2) is to compare CT and ST results with test results of the PST and ECT to explore whether *Sudden* fracture character results in small-scale tests correlate with propagation propensity indicated by the ECT and PST.

2 Methodology

2.1 Data collection

The data collection for this study took place at Mike Wiegele Helicopter Skiing (MWHS) during the winter season 2019/20. The MWHS tenure covers over 4800 km^2 in the Cariboo and Monashee mountains in interior British Columbia, Canada.

Two datasets were acquired in order to answer the research questions in place.

(1) On the one hand the field and snowpack observations of the guides at MWHS were documented which resulted in *Dataset 1*. Thereby the guides were asked to conduct STs and CTs, beside each other at every study site. The MWHS operational stability rating was used to determine how different testing methods represent instable conditions. In addition, the hand


Figure 2: Sample profile site in the Cariboo Mountains

hardness of the layers over- and underlaying the weak layer was documented as well as the weak layer grain type. The documentation further includes the depth and the thickness of the weak layer as well as site characteristics including elevation, exposition and inclination.

(2) In order to evaluate CT and ST results with regard to their informative value on propagation propensity, *Dataset 2* was acquired in a research setting including PSTs and ECTs. Hereby, a research team conducted 3 repetitions of the 4 testing methods being examined
(ST, CT, ECT and PST) at each study location.

In addition to the test results, the hand hardness and grain type as well as size of the layers over- and underlaying the weak layers were recorded. The documentation further includes the depth and the thickness of the weak layer as well as site characteristics including elevation, exposition and inclination. Yellow flags (Jamieson and Schweizer 2005) were used to define weak layer slab combinations that indicate unstable conditions at the profile location.

2.2 Definitions

In the following the parameters used in the analysis of the two datasets are explained.

2.2.1 Test methods

For our comparison we only included test results which are in compliance with testing procedures and methods described in the OGRS (Observation Guidelines and Recording Standards for Weather, Snowpack and Avalanches; CAA, 2016). In the following the techniques and dimensions of the four stability tests used in this analysis are described.

The small- scale tests examined in this paper provide two result components: fracture initiation and fracture character. To obtain this information a small area of the snowpack is isolated, and stress is applied to the relevant weak layer(s) until a fracture occurs.

Compression test (CT)

To conduct a CT, a column with 30 x 30 cm is exposed and stress is induced by tapping on a shovel blade positioned on top the column with increasing force *(Figure 3).* The number of taps needed to initiate a fracture and the characteristics of the fracture compose the CT result.

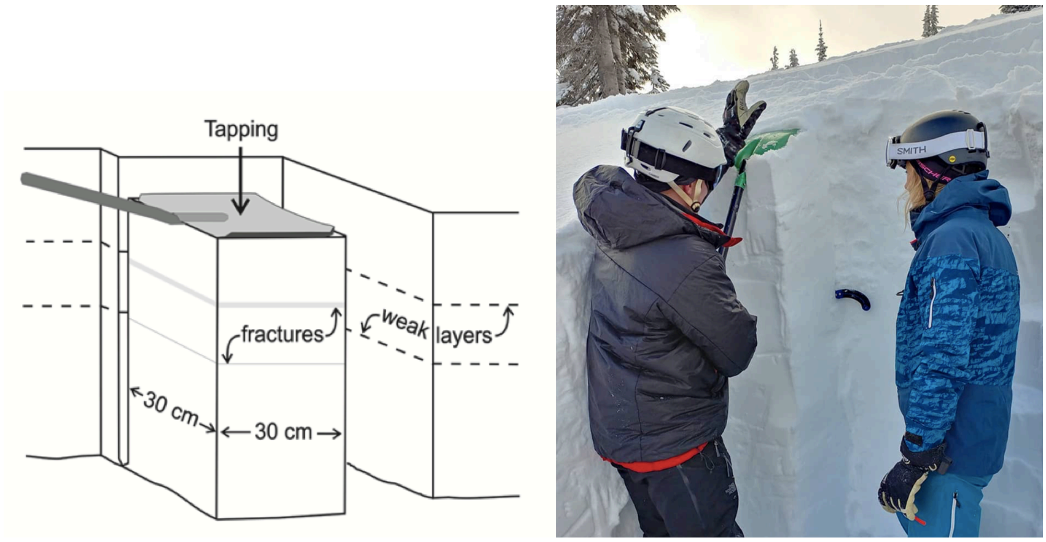


Figure 3: CT technique and dimensions (CAA 2016)

Shovel test (ST)

When conducting a ST, a column with 25 x 35 cm is exposed on all sides with exception of the back wall. Provided the weak layer that should be tested is known, the backside of the column is exposed down to 5 cm below the weak layer. A wedge is removed to ensure the shovel can be easily inserted at the back of the column. With the shovel placed slightly above the weak layer (Figure 4) force is induced by pulling forward on the shovel blade with increasing strength until a fracture occurs in the weak layer. Depending on how much pulling- force is needed until a fracture occurs, the fracture initiation result is then categorized between a very easy to very hard pull. The result is compiled by the force needed to initiate a fracture and the character of the fracture.

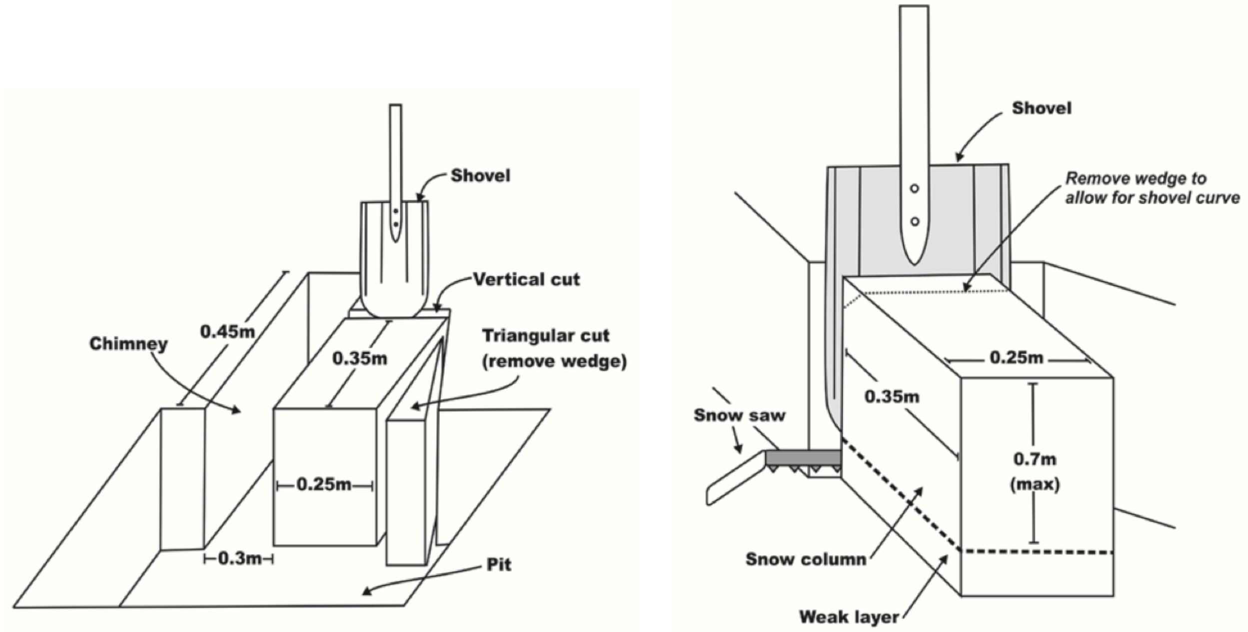


Figure 4: ST technique and dimensions (CAA 2016)

<u>Extended column test (ECT)</u>

The ECT results include information on the fracture initiation as well as its propagation propensity. For the ECT a column with 30 x 90 cm is exposed and the shovel blade is located on the snow surface on one side of the block (Figure 5). Equal to the CT the shovel blade is tapped with increasing force until a fracture occurs. The occurrence of a fracture provides information on the fracture initiation. Secondly, the fracture is observed to determine whether or not it propagates to the end of the column or arrests to determine propagation propensity.

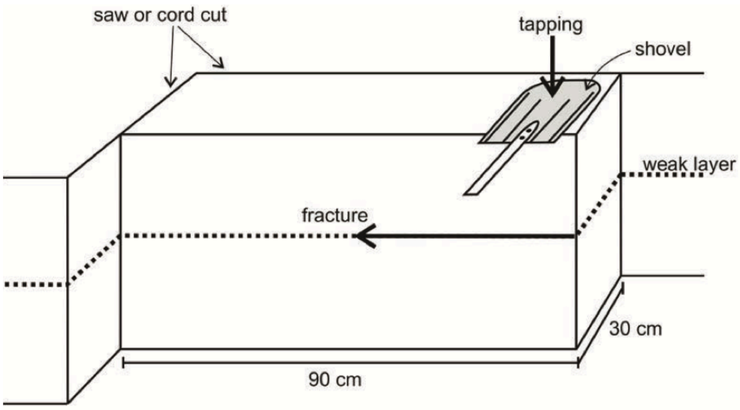


Figure 5: ECT technique and dimensions (CAA 2016)

Propagation saw test (PST)

The PST was specifically designed to determine fracture propagation propensity (Gauthier and Jamieson, 2008). The fracture is initiated by pulling the dull side of a saw through the weak layer until the fracture produced by penetrating the saw reaches a critical length and rapid propagation is initiated (Figure 6). The length of the crack produced by the saw until propagation occurs (x) and the propagation characteristics are described in the results.

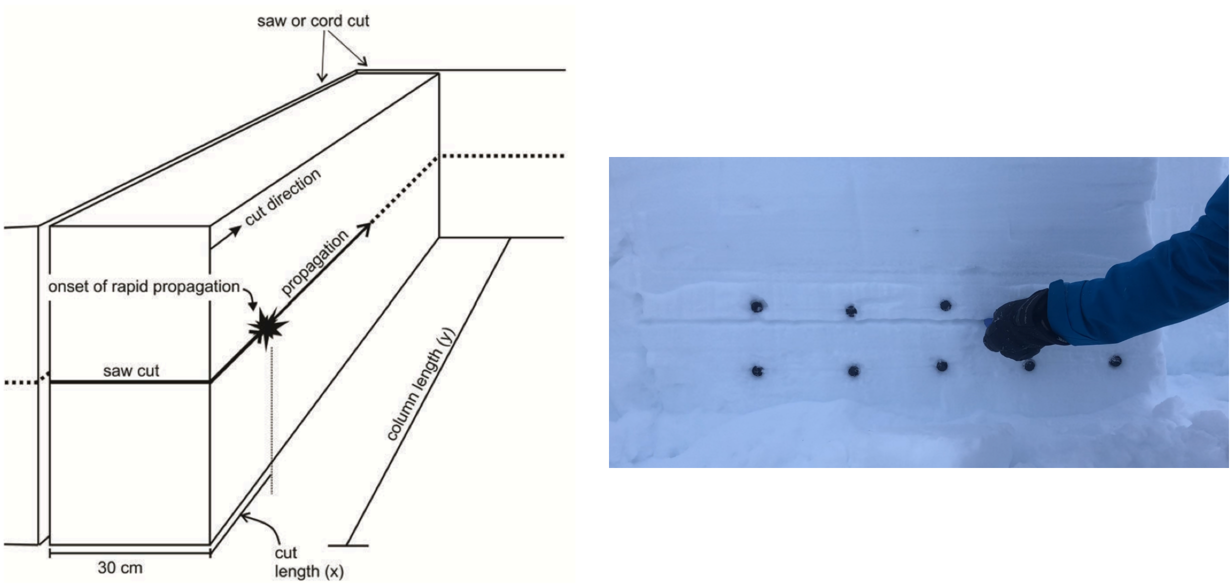


Figure 6: PST technique and dimensions (CAA 2016)

The dimensions of the column are 100 cm in length (y), parallel to the direction of the slope, and 30 cm wide. If the weak layer is more than 100 cm below the snow surface, the length (y) of the column equals the depth of the weak layer.

2.2.2 Fracture Initiation

Fracture initiation occurs if the shear and compression stress applied to the weak layer overcomes its strength (Mccammon and Sharaf, 2005).

In the OGRS, CT and ST fracture initiation results are divided into 5 categories. At MWHS ST fracture initiation results are divided into 7 instead of the 5 categories stated in the OGRS. In order to allow better comparability to other testing methods the 7 categories were compromised into the 5-category rating scale (Table 1).

MWHS Rating System	OGRS Rating System
1. Very Easy	Very Easy
2. Easy	Easy
3. Easy-Moderate	Moderate
4. Moderate	
5. Moderate-Hard	Hard
6. Hard	
7. Very Hard	No Fracture/Very Hard

Table 1: Modification MWHS Rating System to ORGS Rating System

To compare the CT and ST with the ECT and PST, fracture initiation results were divided into the same categories *(Table 1)*. The CTs and ECTs are categorized based on the number of taps while the PST is categorized based on the percentage of the column that was cut before a fracture occurred *(Table 2)*.

Test Method x Result	CT (taps)	ECT (taps)	PST (length of column in %)
Very Easy	< 1 tap	< 1 tap	< 1 %
Easy	1-10 taps	1-10 taps	1-25 %
Moderate	11-20 taps	11-20 taps	26-50 %
Hard	21-30 taps	21-30 taps	51- 99 %
No Fracture	>30 taps	>30 taps	100 %

Table 2: Fracture initiation classification for the CT, ECT and PST.

2.2.3 Fracture character

The fracture character is a parameter to describe the quality of the fracture which is influenced by weak layer properties as well as the characteristics of its adjacent layers (grain size, grain type, hardness). In accordance with the OGRS (CAA 2016) the fracture characteristics of the CT and ST are divided into 4 categories *(Table 3).*

Categories	Sub-categories
1. Sudden	Sudden Planar
	Sudden Collapse
2. Resistant	Resistant Planar
	Progressive Collapse
3. Broken	Broken
4. No Fracture	No fracture

Table 3: Fracture character classification for the CT and ST (CAA 2016).

ECTs are commonly categorized in ECTNs (fracture can be produced but does not propagate to the end of the column), ECTPs (ECT propagating to the end of the column either on tap initiating the fracture or one additional tap) and ECTXs (no fracture was initiated). The PST is divided into results propagating to the end of the column (End), results that lead to a fracturing of the overlying slab (SF) and into fractures self- arresting (Arr) before the end of the column. To allow comparability between the tests, the fracture character results were divided into the same categories as the CT and ST *(Table 4).*

Test Method x Fracture Character	CT & ST	ECT	PST
Sudden	• Sudden Planar • Sudden Collapse	ECTP	End
Resistant	• Resistant Planar • Progressive Collapse	ECTN	• SF • Arr
Broken	Broken	-	-
No Fracture	No Fracture	No Fracture	No Fracture

Table 4: Fracture characteristic classification for the CT, ECT and PST.

2.2.4 Weak layer and interface properties

To examine whether differences can be found between the testing methods which depend on weak layer properties the following weak layer variables were included: Grain type, grain size and hardness. Furthermore, the following interface properties (differences between the weak layer and the overlying and the underlying layer) were examined: Difference in hardness, difference in grain size and the depth of the weak layer.

For *Dataset 1* grain sizes were rarely documented because of the time restriction during guiding operations.

2.2.4.1 Grain type

The grain type categories comprise surface hoar (SH), facets (FC), facet on crust combination (FCxr), precipitation particles (PP), decomposing fragments (DF) and graupel (PPgp).

2.2.4.2 Grain size

The average grain size of the weak layer as well as the layer above and below it were documented in *Dataset 2.* The weak layer grain sizes were divided into the following categories: 0-1 mm, 1-1.5 mm, 1.5-2 mm, 2-3 mm, 3-4 mm and larger than 4 mm.

2.2.4.3 Weak layer hardness

The hardness of the weak layer was measured in hand hardness and divided into the following categories (CAA 2016):

- 1 (F- Fist)
- 2 (4F- Four fingers)
- 3 (1F- 1 finger)
- 4 (P- Pencil)

5 (K- Knife)

2.2.4.4 Difference in hardness

The hardness difference is defined by the difference between the average hardness of the layer overlying the weak layer and the hardness of the weak layer.

2.2.4.5 Difference in grain size

In order to calculate the grain size difference of the interface the average grain size of the layer overlying the weak layer was subtracted from the average grain size of the weak layer.

2.2.4.6 Depth of the weak layer

As found by Schweizer and Jamieson (2007) the critical range for weak layers for human triggered avalanches lies at 18-94 cm below the snow surface. Therefore, all weak layers with less than 19 cm of depth were excluded from this study. The depth of the weak layers was subdivided into 20 cm categories ranging from 20-40 cm of depth to 120 cm and deeper.

2.2.5 Stability Rating

The stability assessment for each specific test location reflects the lead guides (guides with at least 4 years of experience in the area and advanced snow safety education) stability rating for the respective area on a given day. The MWHS tenure is divided into zones, averaging 200 km^2, in order to provide a more accurate stability rating for the different drainages or regions. In each zone skier testing in combination with field observations and multiple snow profile test results lead to the stability rating. The rating system ranges from very poor (1) to very good (7) with 7 representing the most stable conditions and 1 the most unstable (*Table 5*).

Stability Rating	Description
1	Very poor: Snowpack is unstable
2	Poor: Snowpack is rather unstable
3	Poor-Fair: Snowpack is variable (Natural activity has diminished)
4	Fair: Snowpack varies considerably with terrain often resulting in local unstable areas
5	Fair-Good: Snowpack is mostly stable
6	Good: Snowpack is stable
7	Very Good: Very stable snowpack

Table 5: Mike Wiegele Helicopter Skiing (MWHS) Stability Rating System.

2.2.6 Yellow Flags

Jamieson and Schweizer (2005) determined certain weak layer and interface properties (Yellow flags; Table 6) that can be used to identify critical layers.

Layer properties	Critical range
Average weak layer grain size	>1mm
Weak layer Hardness	<3
Grain type	SH, FC or DH
Interface properties	
Difference in grain size	> 0.5mm
Difference in Hardness	>1
Depth of interface	20-80 cm

Table 6: Criteria (yellow flags) for identifying potential weak layers. The hand hardness F, 4F, 1F, P, K is assigned values of 1, 2, 3, 4, 5, respectively. Fractional values are allowed, e.g. 4F+ and 1F- are 2.3 and 2.7.

If 5 or 6 yellow flags are detected in the weak layer slab combination, the slope has a 67-75% chance of being skier triggered with sufficient inclination. If there is 4 or less yellow flags detected there is 59% to 66% chance that it will not be skier triggered (Jamieson and Schweizer, 2005).

2.3 Statistical Analysis

The data being analyzed in this paper is categorical. To test the relationship between two variables, the Pearson's Chi-Square test was used (Pearson, 1900). Hereby frequencies observed in certain categories are compared to frequencies that appear by chance (Field, Miles, and Field, 2012).

As basis for the Chi-Square Test a contingency table with the two variables whose relationship is tested is produced. To test the significance of the results provided in this analysis the probability value (p-value) was used. All results with a p- value smaller than 0.05 were classified as significant. A p-value of 0.05 indicates there is a 5 % chance there is no correlation between the variables tested (Null- Hypothesis) and therefore a value lower than this threshold provides evidence the assumed alternative hypothesis is true. Sankey diagrams, correlation tables and bar charts are used to visualize the data.

In the first part of the analysis *research objective 1* was explored on the basis of *Dataset 1.* Hereby, the fracture initiation and fracture character results of both tests were compared regarding different stability levels as well as regarding various weak layer and snowpack characteristics.

Secondly, *Dataset 2* was analyzed to explore *research question 2*. All testing methods documented in *Dataset 2* (CT, ST, ECT and PST) were compared regarding relevant weak layer and interface properties (yellow flags) with the focus on determining the predictive value of the CT and ST fracture characters regarding propagation propensity. If either of the propagation propensity indication tests (ECT or PST) produced a *Sudden* result in the comparison, propagation propensity was assumed.

3 Results

3.1 Overview

Dataset 1 consists of 298 snow profiles each including a ST and a CT side by side. In some of the profiles more than one weak layer was tested which resulted in 633 test comparisons in total. From the 288 weak layers tested in different locations and days, 5 % of the weak layer were tested under very poor stability, 14 % under poor, 20 % under poor to fair, 34 % under fair, 16 % under fair-good, and 11 % under good stability. The conditions were never rated as very stable in the 2019/2020 season.

In *Dataset 2* 37 profiles are documented. At each profile site 3 CTs, 3 STs, 3 Extended Column Tests (ECTs) and 3 Propagation Saw Tests (PSTs) (on persistent weak layers only) per weak layer were conducted resulting in 138 test comparisons.

3.2 Research objective 1

3.2.1 Stability Rating

Out of 576 CT and ST fracture initiation results (number of taps or respectively shear strength), both test methods show a significant correlation with the stability rating (ST: p-value=2.417e-16; CT: p-value = 0.004642) (Figure 7). Out of all weak layers tested the ST is 7-15 % more likely to produce very easy (VE) to easy (E) results in the very poor (1) to poor-fair (3) stability range than the CT.

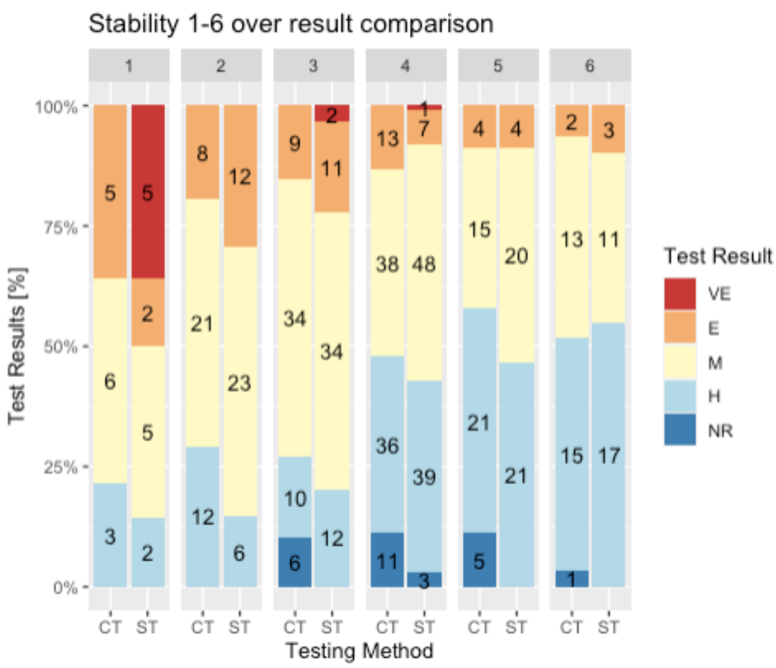


Figure 7: ST and CT test results compared over stability rating from 1-6.

There is a total of 397 fracture character results recorded (*Figure 8*). Out of 88 *Sudden* results produced by the CT 8 % occurred in very poor stability, 26 % in poor, 28 % in poor to fair as well as in fair stability, 8 % in fair to good and 1 % was documented under good stability. Regarding *Sudden* results the two testing methods show a strong agreement among each other. While the ST is twice (47%) as likely to produce *Broken* results, the CT is twice (56 %) as likely to produce *Resistant* results.

The comparison between the fracture character results of both testing methods with the stability rating *(Figure 8)* shows a significant correlation (CT: p-value = $1.78e^{-08}$; ST: p-value = $8.247e^{-08}$).

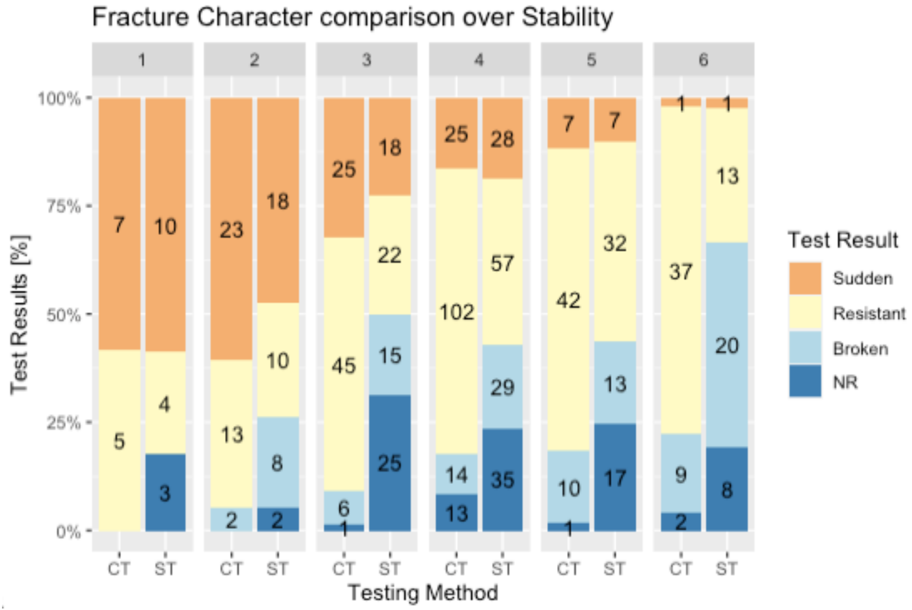


Figure 8: ST and CT fracture character compared over stability rating from 1-6.

3.2.2 Weak layer depth

Further significant results are found when comparing the depth of the weak layers and the fracture initiation results of both testing methods (ST: p-value = $2.556e^{-14}$, CT: p-value=$2.2e^{-16}$) *(Figure 9).* From the 289 weak layer depths documented, 25 % are between 20-40 cm deep, 24 % between 40-60 cm, 20% between 60-80 cm, 16 % between 80-100 cm, 9 % between 100-120 cm and 6 % are deeper than 120 cm.

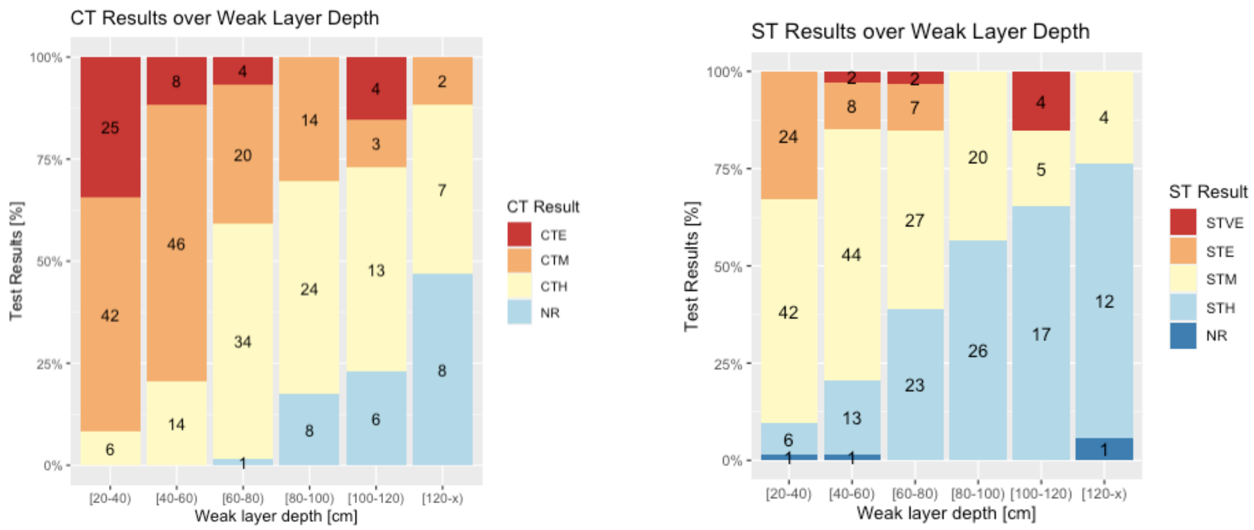


Figure 9: Weak layer depth over test result comparison.

In the depth range between 20-60 cm there is a difference in very easy to easy results between the two tests of 3 %. However, the results become more heterogeneous for the deeper weak layers. On weak layers between 60- 100cm under the snow surface the ST is 15 % more likely to produce a very easy to moderate result than the CT. The percentage of tests during which no fracture could be initiated by the CT (NR) continuously increases in each depth range below 80 cm (80-100 cm: 17%; 100-120 cm: 23%, 120-x: 47%).

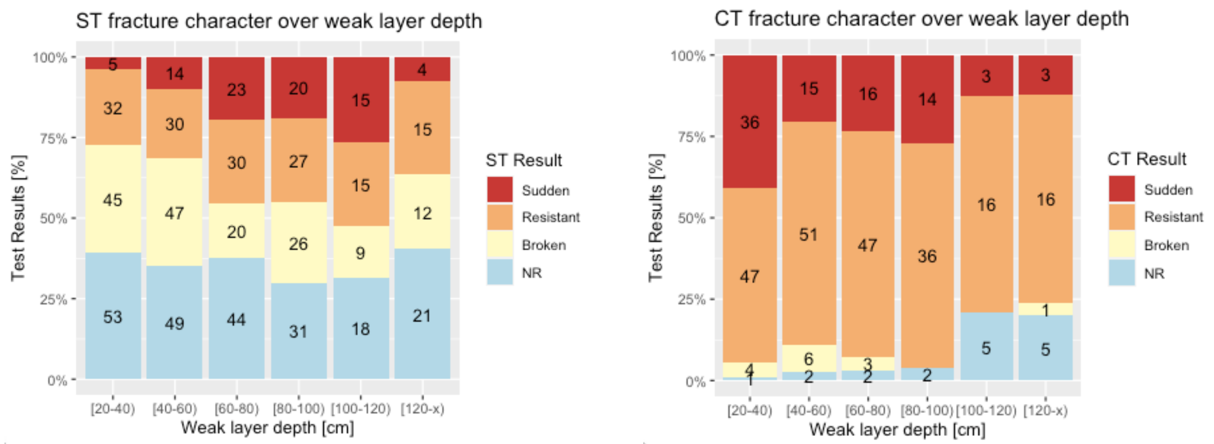


Figure 10: Comparison of fracture characters of weak layer depth.

The comparison of ST and CT fracture character results regarding the depth of the weak layer shows differences. The most obvious difference *(Figure 10)* is that the CT is 39 % more likely to produce *Sudden* results on weak layers less than 60 cm below the snow surface than the ST. In this depth range 31 % of all CTs produce a *Sudden* result while this applies only to 12 % of all STs in the same depth range resulting in 2.4 times as many *Sudden* CT results as *Sudden*

ST results. Between 60-100 cm the ST is 4 % more likely to produce a *Sudden* result than the ST and below 100 cm the ST is twice as likely to produce a *Sudden* result than the CT.

3.2.3 Weak layer grain size and type

The next weak layer characteristics analyzed are weak layer grain sizes and type in comparison to the fracture initiation results. Out of a total of 137 ST and CT fracture initiation results for which the weak layer grain size was documented 38 % of the average grain sizes are above 4 mm, 10 % are between 3 and 4 mm, 16 % are between 2 and 3 mm, 10 % are between 1.5 and 2 mm, 9 % between 1 and 1.5 mm and 18 % are between 0 and 1 mm.

The weak layer grain type was documented in all 289 ST and CT tests from which 33 % of all results occurred on surface hoar (SH) layers, 23 % on facets or facet on crust combinations (FC/FCxr), 19 % on decomposing fragments (DF), 18 % on precipitation particles (PP) and 7 % on graupel (PPgp).

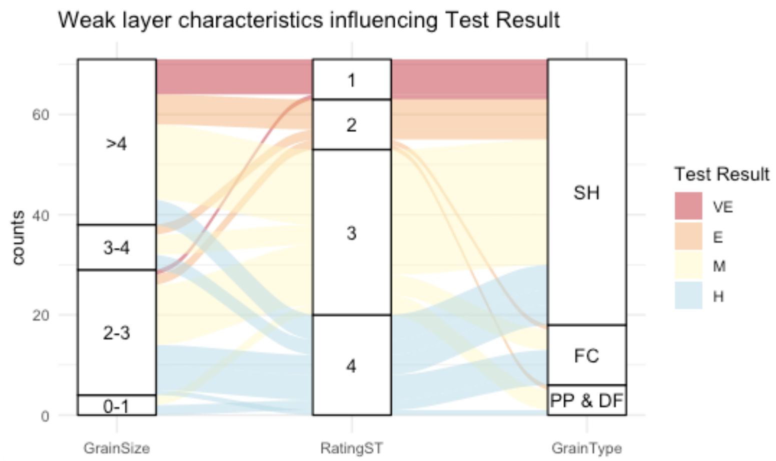


Figure 11: Correlation of weak layer grain size (mm) and grain type with ST Rating.

A significant correlation (p=<0,01) can be found between the average weak layer grain size and both (ST and CT) fracture initiation results. The most significant correlation between very easy and easy test results and surface hoar can be found in weak layers with crystals larger than 4 mm.

Regarding the ST *(Figure 11)* 45 % of all very easy results and easy results (47 in total) are ascribed to SH. 65 % of all very easy and easy results (26 results total) are affiliated with grain sizes over 4 mm. This correlation decreases exponentially with decreasing grain size. No very or easy results were found on weak layers with a grain size of less than 1 mm.

When looking at the correlation between the CT fracture initiation results and the weak layer grain type *(Figure 12)* 34 % of all *easy* results (41 in total) were produced on SH and 52 % of all easy results (21 in total) are affiliated with weak layer grain sizes above 4 mm.

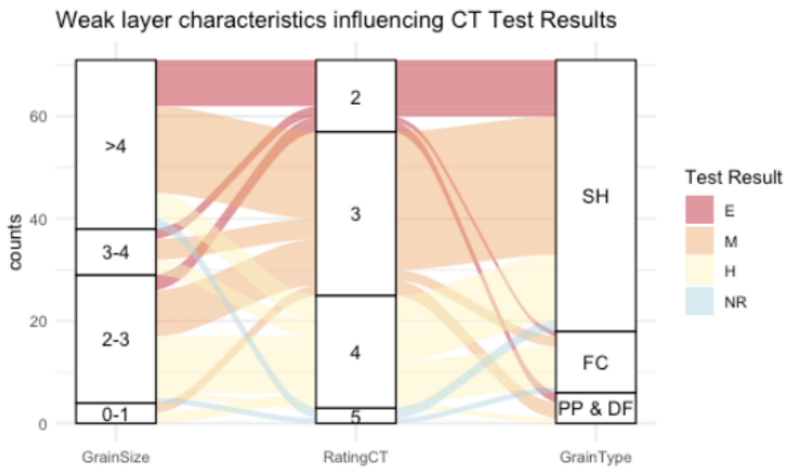


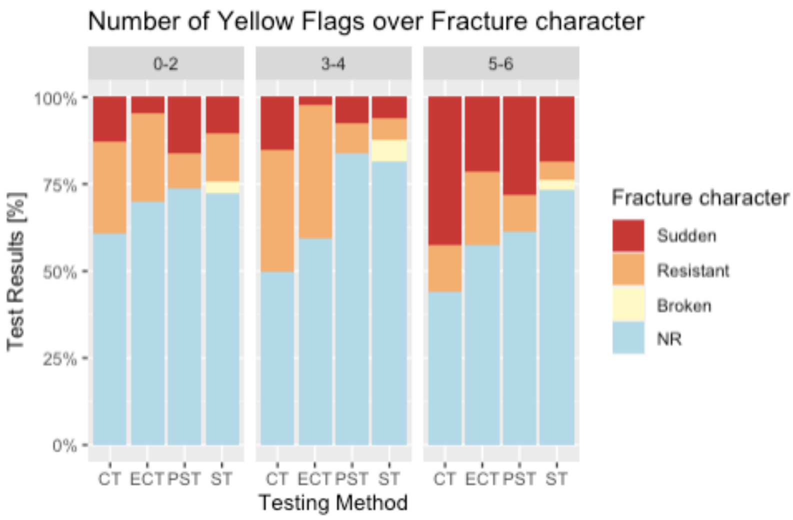
Figure 12: Correlation of weak layer grain size (mm) and grain type with CT Rating.

More than half (CT: 57 %, ST: 56 %) of the fracture initiations were produced on non-persistent weak layers.

3.3 Research objective 2

3.3.1 Yellow flags

The number of yellow flags present in a weak layer/slab combination has a significant correlation ($p < 0.05$) with the fracture character of all results with the exception of the ST ($p=0.28$). There is a total of 333 fracture character results per testing method. For this analysis all layer and interface characteristics that do not represent a yellow flag were excluded. The CT, ST and ECT show an increase in *Sudden* results in each yellow flag category *(Figure 13).* However, while the number of *Sudden* results produced by CTs and ECTs double proportionally for each yellow flag category *(Table 7)* there is only a 1 % increase in *Sudden* results produced by the ST between the 3-4 and 5-6 flags categories. The PST produces more *Sudden* results if 0-2 flags are present than if 3-4 flags are present, but most PST results are produced on layers with 5-6 yellow flags.



Yellow Flag count/ Sudden results	0-2	3-4	5-6
CT	8%	16%	30%
ST	4%	10%	11%
ECT	0%	5%	12%
PST	13%	9%	22%
Total	24	174	135

Table 7: Percentage of Sudden results for each Yellow Flag category.

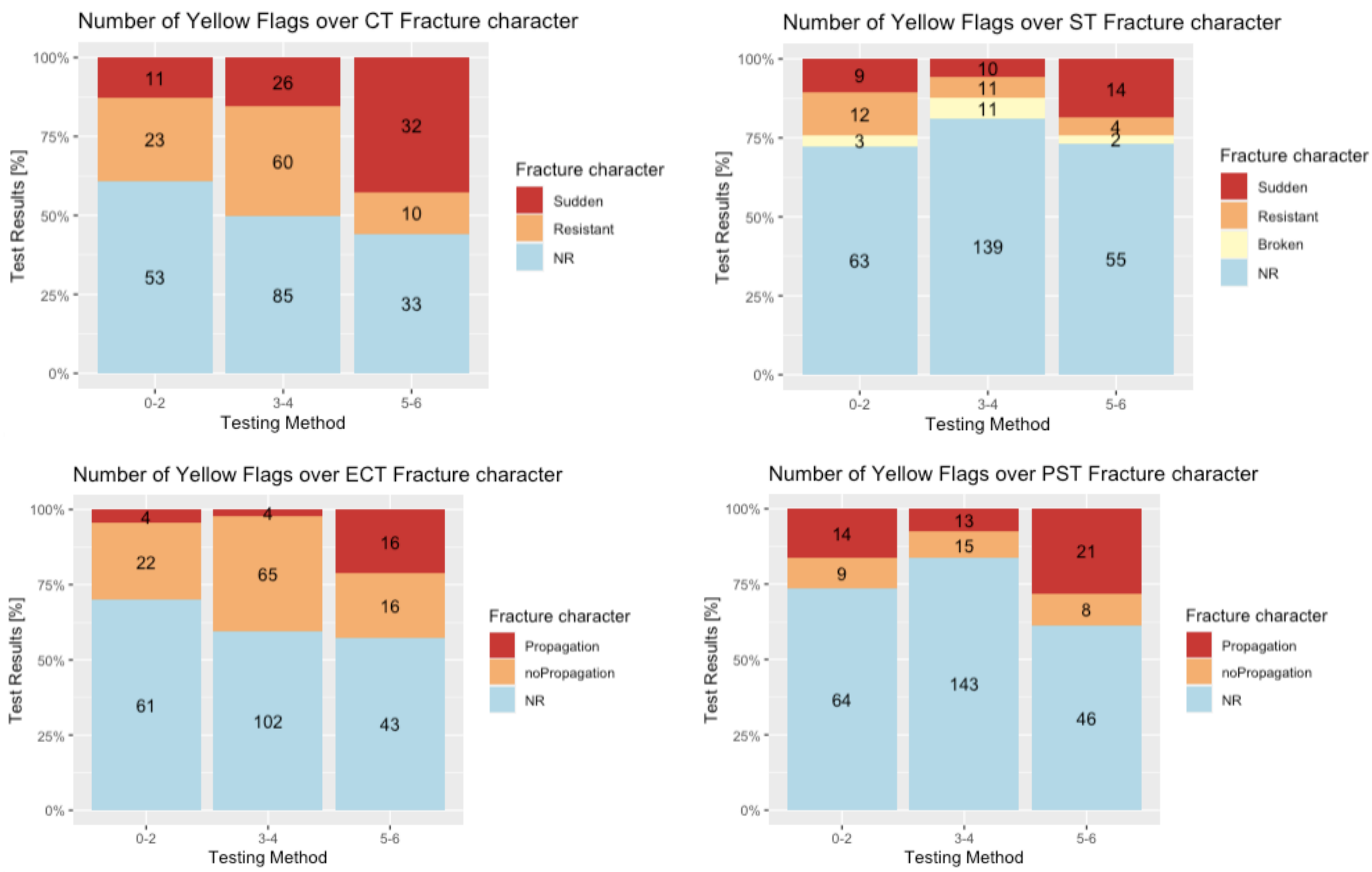


Figure 13: Number of yellow flags over fracture character.

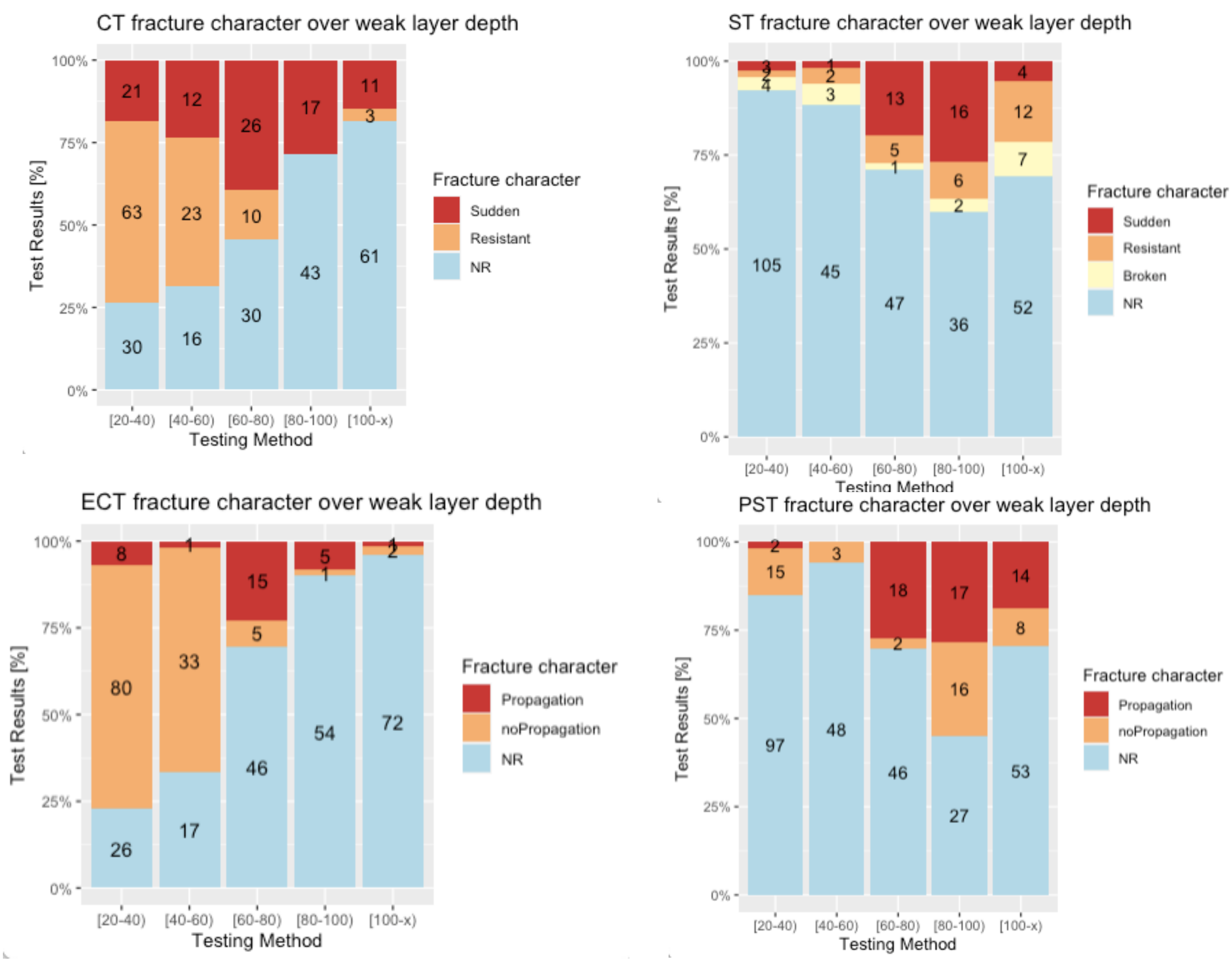
3.3.2 Weak layer depth

The depth of the weak layer significantly influences the results of the different testing methods. At a depth of 20-40 cm the ST fails to produce *Sudden* results on 63% of the weak layer slab combinations with propagation propensity as indicated with the ECT while the CT is 2.6 times as likely to produce *Sudden* results on weak layer/slab combinations than the ECT. Between 40-60 cm there is not enough data available for comparison with 12 *Sudden* CTs and only 1 *Sudden* Result in both, the ECT and the ST.

As indicated in Figure 14 all test methods converge in the depth range between 60-80 cm. 15 ECTPs are produced in this depth range, 13 *Sudden* ST results, 26 *Sudden* CT results and 18 PST results propagate to the end of the column.

Below 80 cm the number of tests initiated by compression (CT, ECT) decline in comparison with the ST and PST with only 6 ECTPs below 80 cm compared to 31 PSTs propagating to the end. Between 80 and 100 cm the PST produces 17 results that propagate to the end of the column which is the same amount as Sudden CT results and the ST produces 16 *Sudden* results in the same depth range.

Between 100 and 120 cm the PST propagates in 11 of the tests which is identical to the number of the *Sudden* CT results while the ST only produces 4 *Sudden* results. Below 120 cm no *Sudden* results were found with the exception of 3 PSTs propagating to the end of the column.



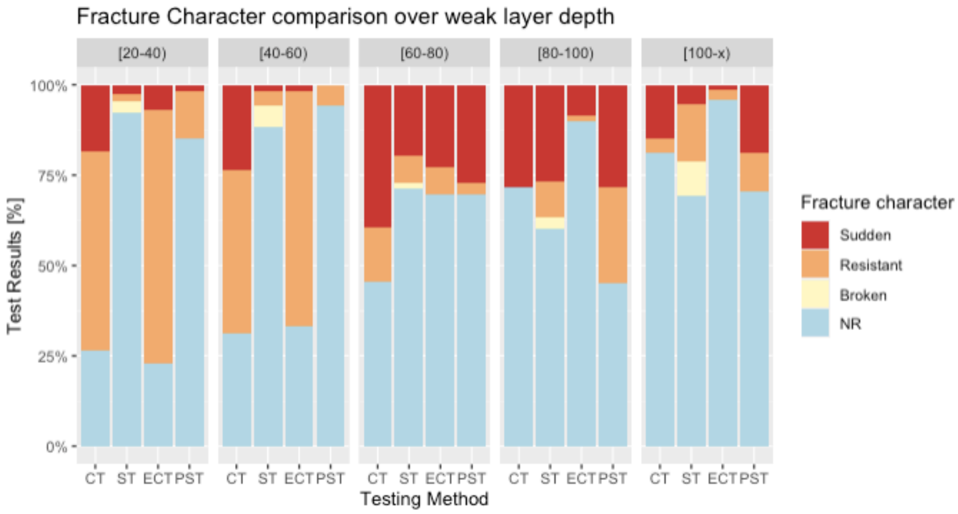
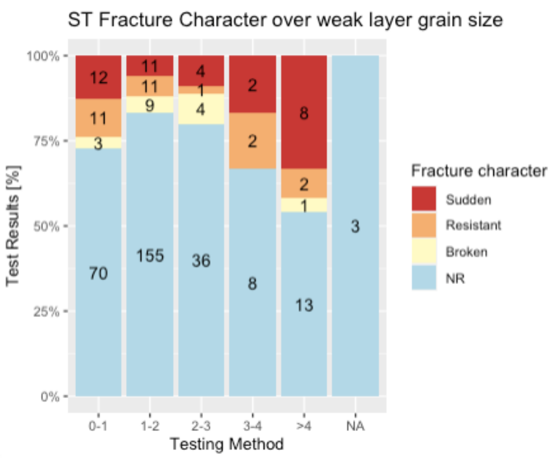
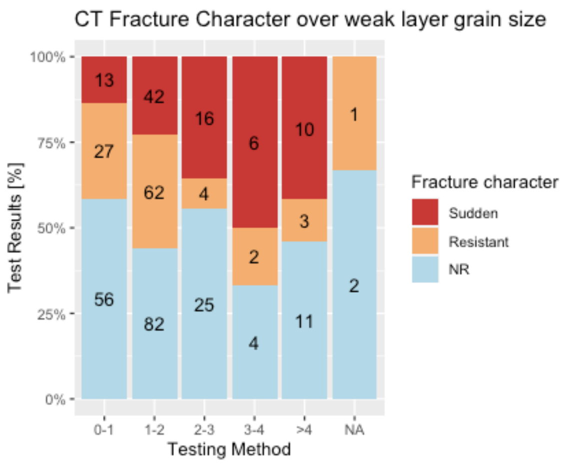


Figure 14: Fracture Character comparison over weak layer depth (cm).

3.3.3 Grain size

There is a significant correlation between the fracture characters of all testing methods and the average grain size of the tested weak layer ($p < 0.05$; *Figure 15).* In the smallest range (0-1 mm) 15 PSTs propagate to the *End* whilst there is 12 *Sudden* ST results and 13 *Sudden* CT results. On weak layers with grain sizes between 1-2 mm 12 ECTs propagate while 42 *Sudden* CT results are produced but only 8 *Sudden* ST results. A similar result is found with weak layer grain sizes between 2- 4 mm (PST *End*: 13, ST *Sudden*: 6, CT *Sudden*: 22). If the average grain size of the layer is larger than 4 mm the testing methods becomes more homogeneous (PST *End*: 12, ST *Sudden*: 8, CT *Sudden*: 10).





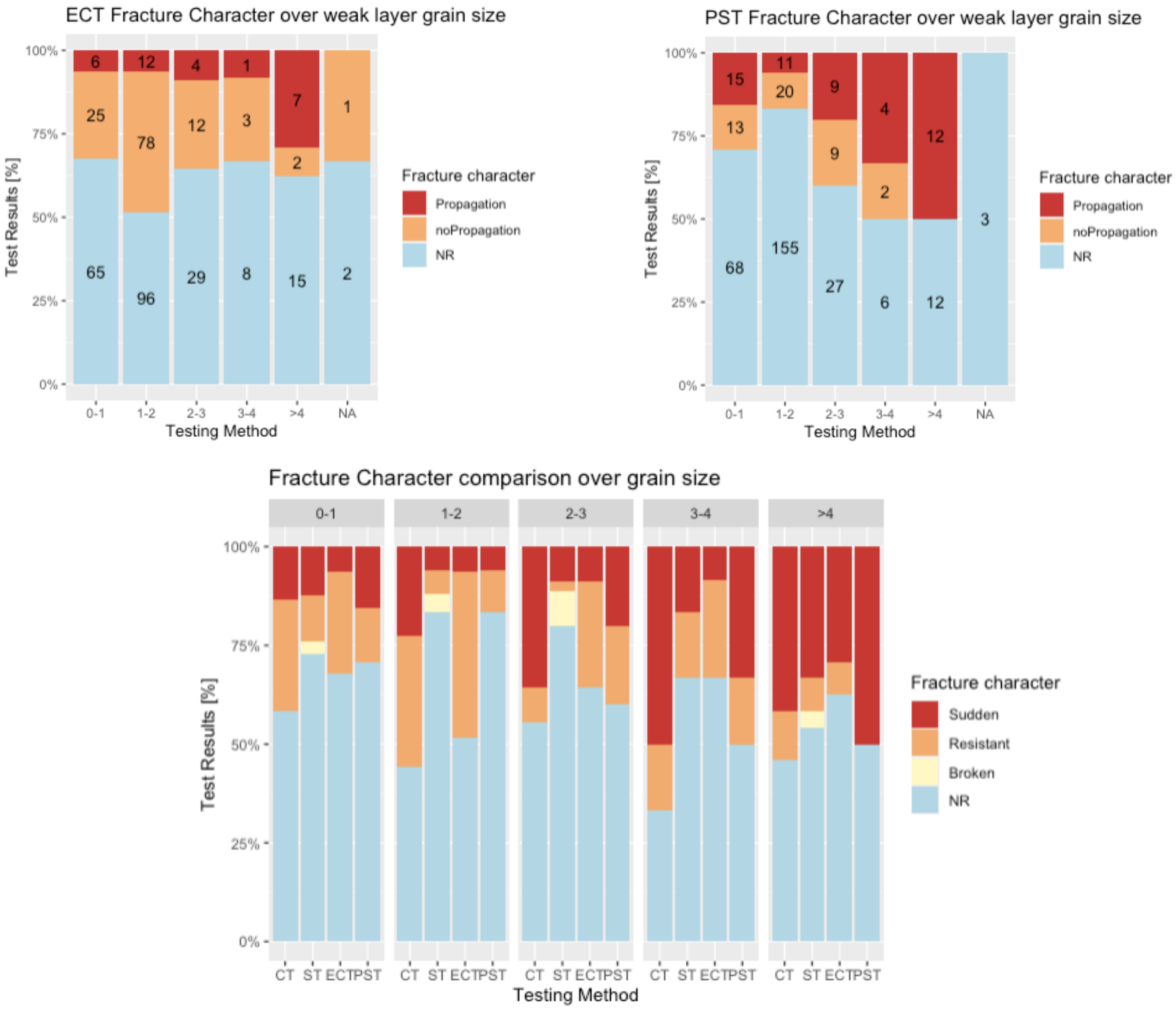


Figure 15: Comparison of fracture character over the average weak layer grain sizes (mm).

3.3.4 Grain Type

The comparison of the weak layer grain type to the fracture character *(Figure 16)* of all tests shows a significant correlation ($p<0.05$). The CT produces more than three times as many (34 in total) *Sudden* results on non- persistent weak layers (PP, DF) than the ECT or the PST. On facetted layers the PST produces 22 *Sudden* results, the CT 20, and the ST 15. The majority of all *Sudden* results documented are documented on SH layers (SH: 89 total; FC: 65 total; DF: 20 total; PP: 31 total). 26 Sudden PSTs compare to 34 *Sudden* CTs and 16 *Sudden* ST on SH layers.

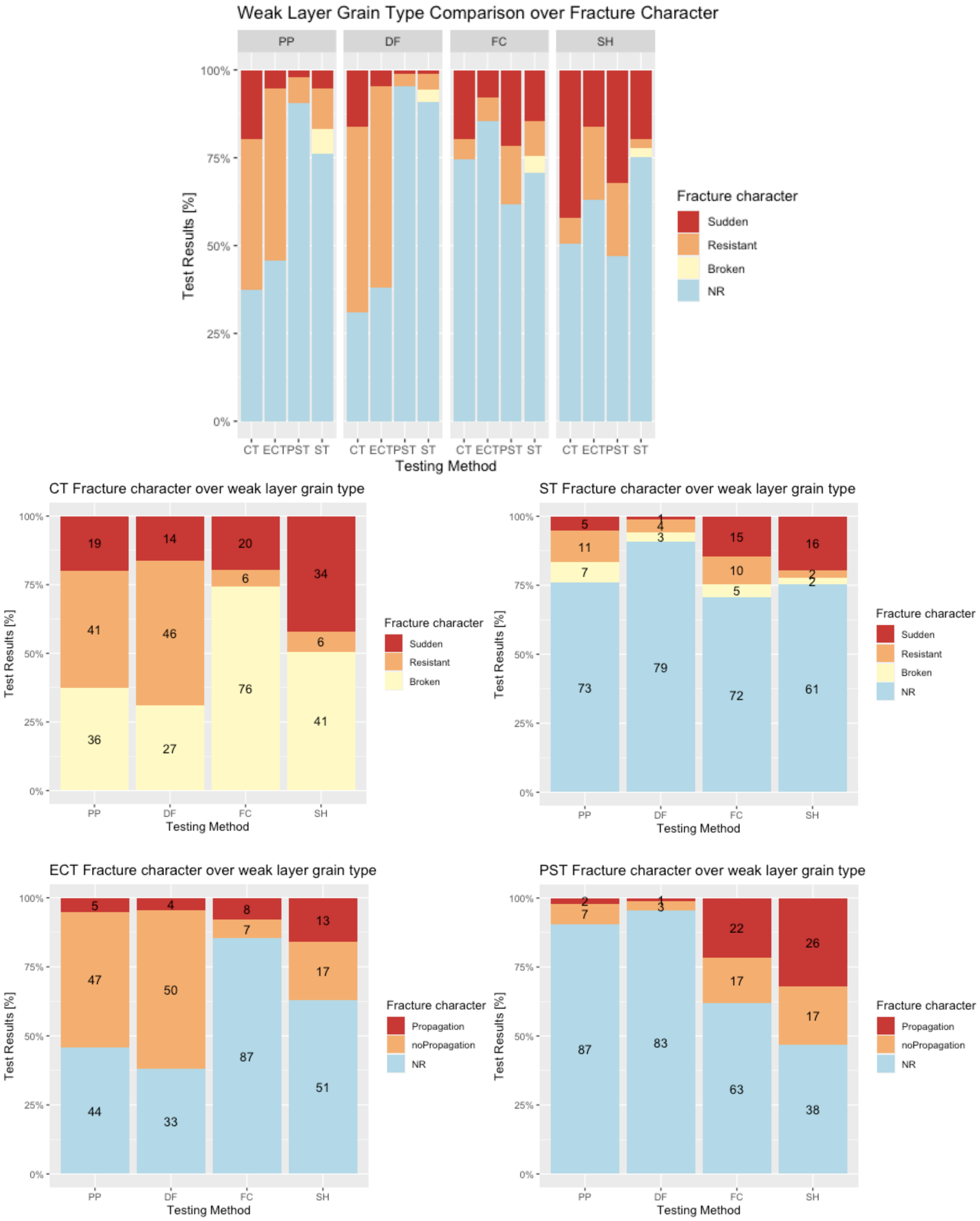


Figure 16: Weak layer Grain Type comparison over fracture character.

3.3.5 Weak layer hardness

The weak layer hardness has a significant correlation with the fracture character of all testing methods ($p < 0.05$). All *Sudden* results are produced on weak layers with a hand hardness of 3 (1F) or less. The most *Sudden* results are found if the weak layer hardness is between 0.5-2 (-F-4F). In this hardness category 88% of all *Sudden* CTs, 67% of all *Sudden* STs, 96% of all

Sudden ECTs and 55% of all *Sudden* PSTs are documented. In the range between 2.5-3 (+4F-1F) the ECT and CT are less likely to produce a *Sudden* result than the ST and the PST.

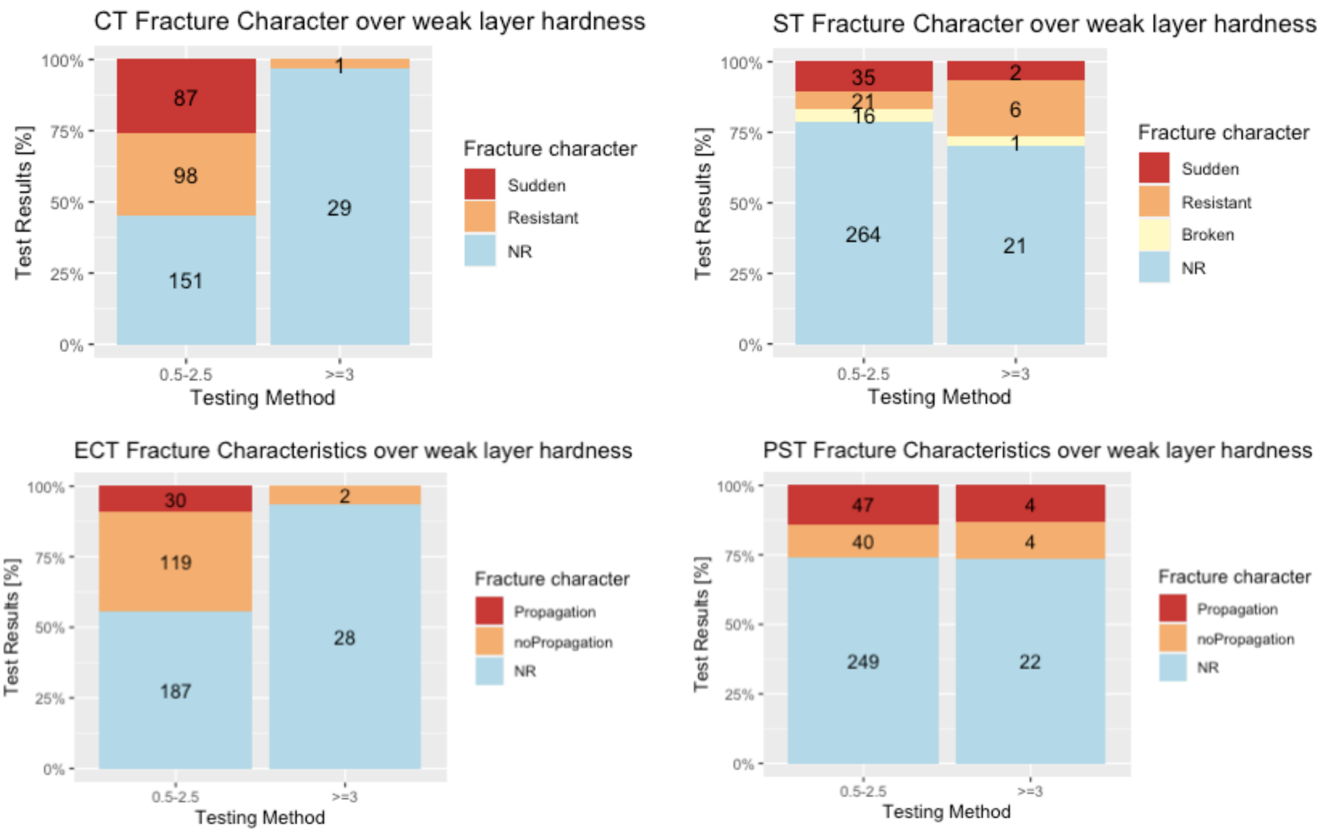


Figure 17: Fracture Character over weak layer hardness

3.3.6 Hardness and Grain Size Difference

The correlation between the Hardness Difference between the weak layer and the overlying slab and the fracture character is significant for all four testing methods ($p < 0.05$). If the hardness difference was more than 1, 55% of the CTs produced a *Sudden* result, 22 % of the STs produced a *Sudden* result, 23 % of the ECTs propagated and 15% of all PSTs propagated to the end of the column. If the hardness difference is less than 1, 15 % of the CTs produced a *Sudden* result, 7 % of the STs produced a Sudden result, 4 % of the ECTs propagated and 14 % of the PSTs propagated to the end of the column. This totals in the following decrease of likelyhood to produce a *Sudden* result if the hardness difference is more than 1 versus less than 1: CT: 40%; ST: 19 %; ECT: 11 %; PST: 1 %.

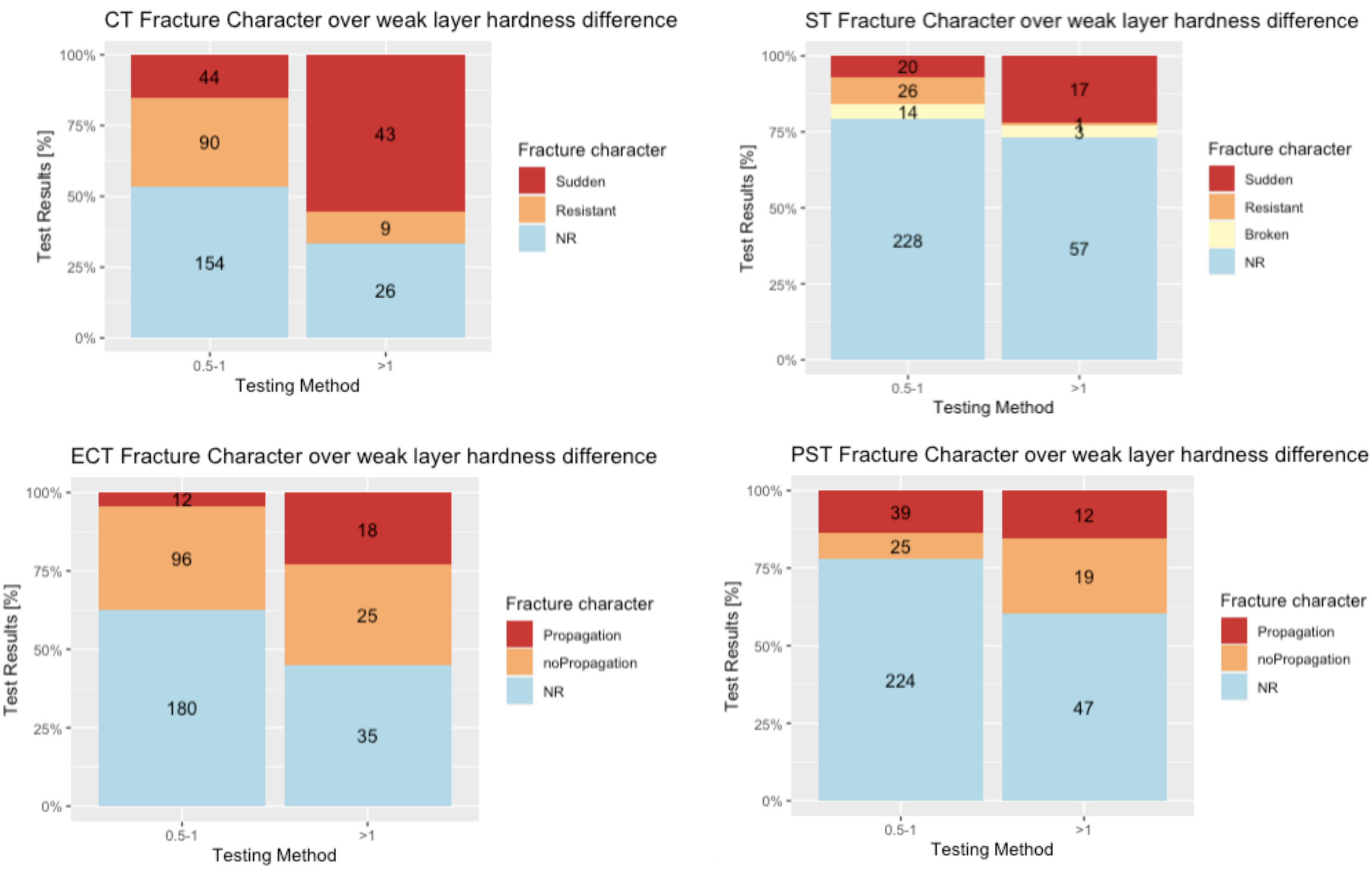


Figure 18: Fracture character over Hardness Difference.

The difference in grain size between the weak layer and the overlying slab correlates significantly with the fracture character of the ECT (p=0,02) and CT (p=0,002) but there is no significant correlation with the fracture characters of the ST (p=0,7) or the PST (p=0,2). The following percentages of all results produced on weak layer slab combinations with a grain size difference of 0.5 mm or more were *Sudden*: ST: 10 %; CT: 24%, PST: 16%, ECT: 8%. For grain size differences of less than 0.5 mm the comparison results in the following percentages: ST: 9 %; CT: 15 %, ECT: 6%; PST: 11 %. Therefore, the likelihood of producing a *Sudden* result for all testing methods decreases if the difference in grain size is less than 0.5 mm. The ST is 1 % less likely to produce a *Sudden* result, the CT is 9 % less likely to produce a *Sudden* result, the ECT is 2 % less likely to produce an ECTP and the PST is 5 % less likely to propagate to the end of the column.

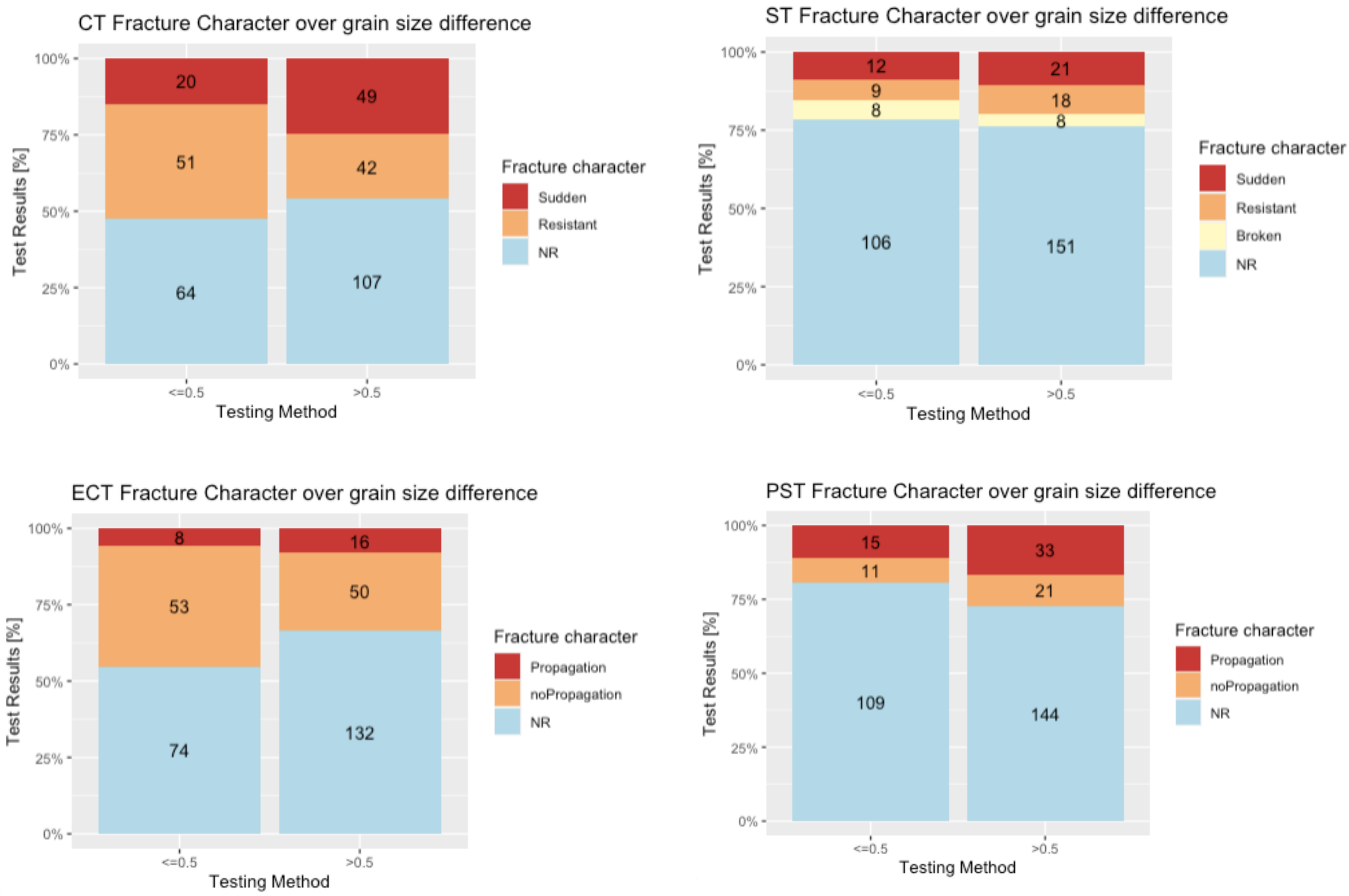


Figure 19: Fracture Character over Grain Size Difference.

4 Discussion

4.1 Research objective 1

To explore *research objective 1* the ST and the CT were compared regarding their predictive value for stability and their dependence on snowpack characteristics. The ST has been repeatedly criticized in past research and reduced to its capability to detect weak layers (Schweizer and Jamieson, 2010) versus a method to significantly support a stability rating. Tremper (1994) states three of the commonly named disadvantages of the ST in his past research: (1) the small sample size, (2) difficulty in interpreting the results and (3) the subjective nature of the test. The CT has the disadvantage of only testing a small sample size but is handled as an objective and easily quantifiable test.

Based on the results from this research both tests are a valid and suitable approach for assessing stability. The fracture initiation and fracture character results of both tests correlate significantly with the stability rating. However, in the range of poor stability (1-3) the results indicate that the ST test initiates fractures easier than the CT. The results of the two testing methods are congruent for *Sudden* results when comparing the fracture character results with the stability rating. However, 2.5 times as many *Sudden* CT results are recorded on weak layers down to 60 cm of depth than *Sudden* ST results. In addition, the CT fracture character result is more likely classified as *Resistant* while the ST results is more often classified as *Broken.* To investigate whether this is due to the triggering mechanism further investigation would be required.

Despite of the limitations and criticism towards both testing methods that were examined the results show strong similarities.

When exploring the dependency of the test results on snowpack characteristics, no significant differences in the two testing methods can be associated with specific snowpack characteristics besides the depth of the weak layer. The CT becomes less sensitive below 60 cm and fails to initiate fractures on some weak layers detected by the ST which are more 80 cm below the snow surface *(Figure 9,10)*. However, the CT produces twice as many *Sudden* results than the ST based on this data which is especially due to the CT producing *Sudden* results on more shallow weak layers down to 60 cm.

Even though a strong agreement is found between the two testing methods in *Dataset 1,* this is not reflected to the same extent in *Dataset 2.* This discrepancy requires further research. A question worth exploring is how objective guides with limited temporal resources are towards test results, especially when conducting multiple testing methods on the same weak layer.

4.2 Research objective 2

The analysis of *research objective 2* was aimed at determining if *Sudden* CT and ST fracture character results correlate with the propagation propensity indicated by the ECT and the PST and how/if this correlation is dependent on weak layer properties.

When focusing only on layers/interfaces with 5-6 yellow flags indicating instable constellations (Jamieson and Schweizer, 2005) the CT produces more than twice as many Sudden results on layers without propagation propensity - as indicated by the ECT and PST - while the ST produces Sudden results on only half of the layers with propagation propensity in this category. While the ECT and CT produce almost twice as many results in each yellow flag category *(Table 7)* the ST does not correlate significantly with the number of yellow flags present. Based on these findings the CT is overly sensitive while the ST fails to detect certain layers with propagation propensity, especially when testing shallow weak layers, non-persistent weak layers and if the hardness difference between the weak layer and the slab is more than 1. Past research (Schweizer and Jamieson, 2000) has found that 52 % of skier triggered avalanches in the Columbia Mountains occurred on non-persistent weak layers or storm snow/old snow interfaces highlighting the relevance of results on non-persistent weak layers. Even though, the tests initiated by compression decrease with the depth of the weak layer in comparison to the other testing methods, the CT is effective in detecting weak layer/slab combinations with propagation propensity down to and below 120 cm when compared with the PST. The ECT on the other hand fails to produce a result indicating propagation on a third (29%) of the weak layer slab combinations with propagation propensity as indicated by the PST at a weak layer depth between 80-100 cm but detects a majority of the weak layer with propagation propensity at skier triggerable depth (Schweizer and Jamieson 2007). The PST produces more results than the other testing methods on small weak layer grain sizes (1 mm or less) and on facet layers. This is likely due to the depth of some of the facet layers tested as the PST also produced the most *Sudden* results on layers below 100 cm under the snow surface. In general, propagation propensity of weak layer/slab combination increases below 60 cm.

The testing methods correlate most significantly on layers that are between 60-80 cm below the snow surface, if the average weak layer grain size is larger than 4mm and if the weak layer hardness is 4 fingers or less.

5 Conclusion and Outlook

To sum up the conclusion for *research question 1* there is a high level of agreement between the ST results and the CT results in this dataset. While the CT is a validated and recognized testing method, the ST is often criticized by researchers and practitioners. On the basis of *Dataset 1* both testing methods are valid due to the high degree of agreement between the ST and the CT that for guides who have a lot of experience in doing STs, the results of the ST are as valuable in gaining information about the stability as CT results.

However, the similarity between the two testing methods indicated in *Dataset 1* is not reflected to the same extent in *Dataset 2.* This discrepancy requires further research. A question worth exploring is how objective guides with limited temporal resources are towards test results, especially when conducting multiple testing methods on the same weak layer.

Based on the results from *research question 2* the CT fracture character is a reliable indicator for propagation propensity. If the weak layer is close to the snow surface (down to 60 cm), the weak layer is non-persistent or if there is a difference in hardness of more than 1 the CT is overly likely to produce Sudden results which means the result should be further investigated to make a judgement on propagation propensity. The characteristics stated above also describe situations in which the ST fails to detect critical layers. Therefore, the ST is an inadequate tool to further investigate the situation. The ST only provides a comparable indication for propagation propensity if the weak layer is well pronounced (e.g. grain size > 4mm) and at a depth between 60-100 cm. Based on the results from this study the ECT fails to detect critical weak layers which are more than 80 cm below the snow surface which indicates that caution must be taken when applying the ECT to test deeper weak layers. Considering the type of problems discovered when applying the CT, the ECT does offer a good basis to further investigate *Sudden* CT results on weak layer slab combinations on which the CT is overly likely to produce a *Sudden* result on. Vice versa the CT is a suitable method to further investigate the snowpack for weak layers more than 80 cm below the snow surface which can likely not be initiated by the ECT. Therefore, it can be concluded that in very critical situations all testing methods examined are capable of detecting weak layers and also give an indication on propagation propensity. If the avalanche problem is less pronounced, non-persistent or close to the surface the assessment of the snowpack requires more than one testing method.

6 References

Birkeland, Karl, and Ron Simenhois. 2006. “The Extended Column Test: A Field Test for Fracture Initiation and Propagation.” *International Snow Science Workshop*. Telluride, CO.

CAA. 2016. *Observation Guidelines and Recording Standards for Weather, Snowpack and Avalanches*. 6.1.

D. M. McClung, and Jürg Schweizer. 2006. “Fracture Toughness of Dry Snow Slab Avalanches from Field Measurements.” *Journal of Geophysical Research* 111. https://doi.org/0.1029/2005JF000403.

Field, Andy, Jeremy Miles, and Zoe Field. 2012. *Discovering Statistics Using R*. SAGE Publications Ltd.

Föhn, Paul. 1986. “The Rutschblock as a Practical Tool for Slope Stability Evaluation.” In *Avalanche Formation, Movement and Effects*.

Gauthier, Dave, and Bruce Jamieson. 2008. “Fracture Propagation Propensity in Relation to Snow Slab Avalanche Release: Validating the Propagation Saw Test.” *Geophysical Research Letters* 35 (13). https://doi.org/10.1029/2008GL034245.

Herwijnen, Alec van, and Bruce Jamieson. 2004. “Fracture Character in Compression Tests.” *International Snow Science Workshop*. Jackson Hole, Wyoming.

Jamieson, Bruce, and Colin Johnston. 1996. “The Compression Test for Snow Stability.” In *Snow Cover Stability, Avalanche Initiation and Forecasting*. Banff, AB.

Jamieson, Bruce, and Cameron Ross. 2012. “The Propagation Saw Test: Slope Scale Validation and Alternative Test Methods.” *Journal of Glaciology* 58 (208). https://doi.org/10.3189/2012JoG11J192.

Jamieson, Bruce, and Jürg Schweizer. 2005. “Using a Checklist to Assess Manual Snow Profiles.” *Avalanche News*.

Johnson, Ron F., and Karl Birkeland. 2002. “Integrating Shear Quality into Stability Test Results.” *International Snow Science Workshop*. Penticton, BC.

Mccammon, Ian, and Don Sharaf. 2005. “Integrating Strength, Energy and Structure into Stability Decisions: So You Dig a Pit and Then What?” *The Avalanche Review* 23 (3): 18–19.

McClung, David, and Peter Schaerer. 2006. *The Avalanche Handbook*. 3rd ed. The Mountaineers Books.

Pearson. 1900. “On the Criterion That a given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling.” *Philosophical Magazine Series*, 302, 5 (50): 157–75.

Ross, Cameron, and Bruce Jamieson. 2008. “Comparing Fracture Propagation Tests and Relating Test Results to Snowpack Characteristics.” *International Snow Science Workshop*. Whistler,BC.

Schaerer, Peter. 1988. “Evaluation of the Shovel Shear Test.” *International Snow Science Workshop*. Whistler,BC.

Schweizer, Jürg, and Bruce Jamieson. 2000. “Field Observations of Skier-Triggered Avalanches.” *International Snow Science Workshop* . Big Sky, Montana.

———. 2007. “A Threshold Sum Approach to Stability Evaluation of Manual Profiles.” *Cold Regions Science and Technology*, no. 47: 50–59. https://doi.org/10.1016/j.coldregions.2006.08.011.

———. 2010. “Snowpack Tests for Assessing Snow-Slope Instability.” *Annals of Glaciology* 51 (54).

Schweizer, Jürg, Ian Mccammon, and Bruce Jamieson. 2008. “Snowpack Observations and Fracture Concepts for Skier-Triggering of Dry-Snow Slab Avalanches.” *Cold Regions Science and Technology* 51 (2–3): 112–21.

Tremper, Bruce. 1994. “Simple Snow Stability Tests.” *The Avalanche Review* 13 (2).
Walcher, Matthias, Pascal Haegeli, and Sven Fuchs. 2019. “Risk of Death and Major Injury from Natural Winter Hazards in Helicopter and Snowcat Skiing in Canada.” *Wilderness & Environmental Medicine*, 1–9.
Winkler, Kurt, and Jürg Schweizer. 2008. “Comparison of Different Snow Stability Tests Including the Extended Column Test.” *International Snow Science Workshop*, 393–400. Whistler, BC.