Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL078069

Key Points:
- Fracture processes relevant for dry snow slab avalanche release were modeled with FE
- Key snow properties control failure initiation, crack propagation, and tensile support
- Validation with independent field data supports the presented three-step snow instability model

Supporting Information:
- Supporting Information S1

Correspondence to:
B. Reuter,
reuter@slf.ch

Citation:
Reuter, B., & Schweizer, J. (2018). Describing snow instability by failure initiation, crack propagation, and slab tensile support. *Geophysical Research Letters*, *45*, 7019–7027. https://doi.org/10.1029/2018GL078069

Received 27 MAR 2018
Accepted 23 MAY 2018
Accepted article online 1 JUN 2018
Published online 26 JUL 2018

Describing Snow Instability by Failure Initiation, Crack Propagation, and Slab Tensile Support

B. Reuter[1,2] **and J. Schweizer**[1]

[1]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland, [2]Department of Civil Engineering, Montana State University, Bozeman, MT, USA

Abstract Snow instability is a generic term describing the propensity of a snow slope to avalanche. In need of a concise mechanics-based concept we suggest a framework based on failure initiation, crack propagation, and slab tensile support. Following these three steps we modeled three metrics from mechanical data, which we derived from snow micropenetrometer signals. Verifying the metrics with field measurements confirmed that slab thickness and weak layer strength typically influence failure initiation, elastic modulus and weak layer fracture energy largely control crack propagation, and slab thickness and tensile strength provide the required tensile support. For all three metrics, considering slab layering was essential. Validation with signs of instability showed that the most accurate model includes all three steps – suggesting that snow instability can be described by failure initiation, crack propagation, and slab tensile support. Further validation is needed to assess the framework's potential for operational use.

Plain Language Summary Snow avalanches threaten people and infrastructure in all snow-covered mountain regions. Avalanche mitigation includes avoiding avalanches in space and time. The latter requires avalanche forecasting, that is, predicting snow instability or the probability of a snowpack to avalanche. As properties of the mountain snowpack vary spatially and evolve with time, modeling approaches are better suited than point observations to assess snow instability. However, for simulating snow instability in snow cover models quantitative measures are needed, preferably reflecting the processes leading to slab avalanche release. Dry snow slab avalanche release can be interpreted as a sequence of fractures, including failure initiation and crack propagation. We follow these steps and complement our recently developed metrics describing failure initiation and crack propagation with a criterion for the tensile support of the slab. We identify the snow properties that are most relevant to drive the three processes and validate the metrics with independent field observations. The validation confirmed that snow instability is best interpreted as a combination of failure initiation, crack propagation, and slab tensile support. Our new model to describe snow instability on mountain slopes is now ready to be assessed in operational use for avalanche forecasting.

1. Introduction

The release of a dry-snow slab avalanche is a fracture mechanical process. Failure initiation in the weak layer below the slab followed by rapid crack propagation across the slope are key processes before snow masses start moving downhill (Schweizer et al., 2003; van Herwijnen & Jamieson, 2007). Whumpfs and shooting cracks may be observed alongside (Jamieson et al., 2009). Snow instability tests allow observing these processes separately (Schweizer & Jamieson, 2010). However, slope stability evaluation from point observations, for example, by stability tests, is hampered by the variable nature of the seasonal snow cover (Birkeland, 2001; Kronholm & Schweizer, 2003).

Hence, estimating snow instability, which is crucial to avalanche danger forecasting and infrastructure risk assessment, builds on record keeping of avalanche activity, signs of instability, and stability test results. This approach is used because presently there is no metric describing snow instability and even less so avalanche danger (Techel & Schweizer, 2017).

Snow cover modeling offers an alternative way to obtain snow stratigraphy information (Bartelt & Lehning, 2002; Brun et al., 1992; Vionnet et al., 2012) with high temporal and spatial resolution. Coupling of such snow cover models with numerical weather prediction models allows predicting snowpack properties in time and space (Bellaire & Jamieson, 2013; Quéno et al., 2016; Vionnet et al., 2016). By adding a mechanical model to this model chain we can reach out for predicting snow instability in time and space (Vernay et al., 2015).

To this end, a description of snow instability at the scale of the snow cover or a point in the terrain is much needed, yet presently lacking. It seems natural that such a description is based on the processes involved in avalanche release, namely, failure initiation and crack propagation, and employs snowpack data acquired with the snow micropenetrometer (SMP; Schneebeli & Johnson, 1998). Only recently, a first framework has been presented describing point snow instability (Reuter, Schweizer, & van Herwijnen, 2015) that was used for spatial (Reuter, van Herwijnen, et al., 2015) and temporal analyses of instability (Schweizer et al., 2016).

However, during the initial phase of dynamic crack propagation the tensile strength of the slab influences how far a running crack will propagate (Gaume et al., 2015). This finding was so far supported by a preliminary analysis of field data showing that denser slabs showed more extensive crack propagation (Schweizer et al., 2014), and seems important for observer-independent snow instability evaluation (Reuter, van Herwijnen, & Schweizer, 2016).

Our aim is to provide a framework for snow instability evaluation from point observations. To this end, we introduce a slab tensile criterion complementing the existing criteria for failure initiation and crack propagation. We will only briefly recall the latter two and then validate all three criteria with field data. Based on the comparisons with field experiments we highlight the main snow cover properties that drive the three processes influencing snow instability. Eventually, we assess the criteria's performance to predict snow instability based on observations of signs of instability. Our findings suggest that a three-step approach performs slightly better than the previously presented two-step approach—emphasizing the role of slab tensile support for crack propagation.

2. Materials and Methods

2.1. Snow Mechanical Properties

Our snow instability metrics are based on snow mechanical properties derived from field measurements for slab layers, the weak layer, and a basal layer. Mechanical properties include snow density, elastic modulus E, tensile strength σ_t, specific fracture energy w_f, and weak layer strength σ_{WL}. Snow density, effective elastic modulus, and weak layer specific fracture energy were derived from the SMP signals as described by Reuter et al. (2018) following the approach for signal interpretation suggested by Löwe and van Herwijnen (2012). We used the density model reported by Proksch et al. (2015), but instead of using their coefficients based on Alpine and Arctic samples, we derived a local calibration, if manual snow density measurements and corresponding reference SMP measurements were available at the study site on that particular day. The tensile strength σ_t was determined after Jamieson and Johnston (1990) from SMP-derived snow density, and the strength of the weak layer σ_{WL} was approximated with the SMP-derived microstructural strength (Johnson & Schneebeli, 1999).

2.2. Modeling Snow Instability

We derived three snow instability metrics, each one based on finite element (FE) modeling, as is usually required when loading a multilayer material. The first two metrics are the failure initiation and crack propagation criteria introduced by Reuter, Schweizer, and van Herwijnen (2015).

To assess the stability at the depth of the weak layer during localized skier loading at the snow surface, we calculated the failure initiation criterion as the ratio of the strength of the weak layer σ_{WL} and the maximum additional shear stress at the depth of the weak layer $\Delta\tau$. The latter was obtained from a FE simulation assuming a static strip load on top of the slab sitting on a weak layer. Alternatively, we calculated the maximum additional shear stress for a strip load from the analytic formulation presented by McClung and Schweizer (1999) using an equivalent slab thickness (Monti et al., 2016), which partly accounts for slab layering.

To assess the propensity for crack propagation in the weak layer below the slab, we calculated the critical crack length r_c based on the work of Heierli et al. (2008) but used the expression for the mechanical energy presented by van Herwijnen et al. (2016) including their calibration for slope angle and ratio of crack length to slab thickness.

To complement the two above-mentioned metrics of snow instability, we introduced a tensile failure criterion T_i. This criterion relates the tensile strength of each slab layer $\sigma_{t,i}$ to the maximum tensile stress in each

layer σ_i at the onset of crack propagation. The slab tensile criterion T is expressed as the thicknesses of the layers h_i with a critical value T_i and normalized by total slab thickness H:

$$T = \frac{\sum_i f_i h_i}{H}, \text{with } f_i = \begin{cases} 1 \text{ if } T_i = \frac{\sigma_{t,i}}{\sigma_i} < 1, \\ 0 \text{ if } T_i = \frac{\sigma_{t,i}}{\sigma_i} \geq 1. \end{cases} \quad (1)$$

With a FE model we simulated the tensile stresses σ_{yy} in the slab at the onset of crack propagation. In the model, the slab is unsupported along the critical crack length and supported by the weak layer from the crack tip upslope (Figure S1 in the supporting information). For each layer we extracted the maximum tensile stress within a band of 30 cm width that is centered at the tensile stress maximum at the snow surface and extends down to the weak layer.

2.3. Statistical Measures

We used the Pearson correlation coefficient r_p, or Spearman's correlation coefficient r_s for ordinal data, and the value of significance p to describe the strength of a linear relationship between two quantities. To verify how the metrics compare with measurements, we used the coefficient of determination R^2 (with the definition of the fraction of unexplained variance), the root-mean-square error (RMSE), and the mean absolute error. To describe classification accuracy, we calculated the probability of detecting unstable cases (POD; also called true positive rate TP), the probability of correctly predicting stable cases (PON), and the false alarm rate (FP = 1 − PON; also called false positive rate). The mean of POD and PON is called unweighted average accuracy (RPC; Fawcett, 2006; Schweizer & Jamieson, 2010; Wilks, 2011).

We graphed model predictions in receiver operating characteristic (ROC) space (Fawcett, 2006), which depicts the tradeoff between hits and false alarms and allows comparing models with different classifiers. The model scores, that is, the probability that an observation belongs to the unstable class, were derived from the probability estimates of a logistic regression analysis or from the class proportions at the leaves of a classification tree. We normalized variables and assumed a binomial distribution in the logistic regression. Classification trees had a maximum of two splits per predictor variable. The Gini coefficient (GINI) is a measure of classification quality and relates to the area between the ROC curve and the 1:1 line. The point on the ROC curve at maximum distance perpendicular from the diagonal corresponds to the probability threshold yielding the highest value of the true skill statistic (TSS = POD − FP, Wilks, 2011).

3. Field Data

From routine measurements and field campaigns during winter seasons between January 2003 and March 2015 we accumulated a field data set containing 139 measurements with the SMP and corresponding observations of snow stratigraphy and snow instability. The data were all collected around Davos (Eastern Swiss Alps).

The SMP measurements contain records of penetration resistance to a depth well below the weak layer that we had identified beforehand in an adjacent manually observed snow profile (Canadian Avalanche Association, 2007); profiles were complemented with snow instability tests. Thus, 77 of the 139 SMP measurements can be related to a propagation saw test (PST; Gauthier & Jamieson, 2008) with column length requirements given in van Herwijnen et al. (2016) and 62 of the 139 SMP measurements to a Rutschblock (RB) test (Föhn, 1987). In 122 of the 139 cases our data also include records of signs of instability, that is, the absence or the presence of whumpfs, shooting cracks, and fresh avalanches (Jamieson et al., 2009) at the field site.

In other words, we deal with three data sets: one with signs of instability (N = 122), a PST (N = 77), and a RB (N = 62) data set (Reuter et al., 2018). While we use the PST and the RB data to verify our metrics, we use the larger data set including signs of instability to assess the prediction performance of combining our three metrics. The PST and the RB data described here include the ones previously used by Reuter, Schweizer, and van Herwijnen (2015) (N = 129 of 139).

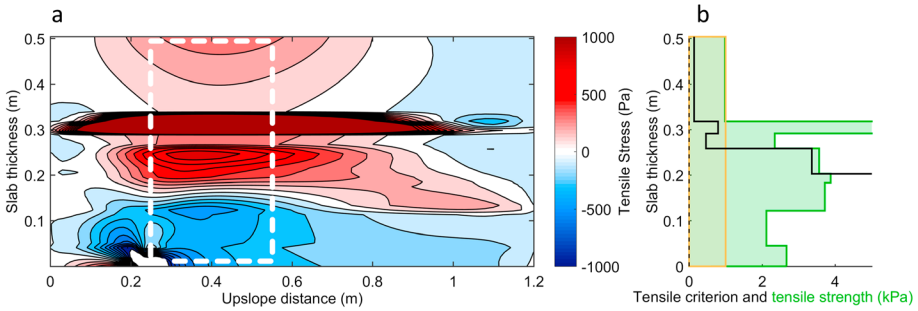

Figure 1. (a) Tensile stresses in the slab featuring a stress concentration around a melt freeze crust modeled with FE for a propagation saw test where the slab fractured at a crack length r_c = 0.23 m. Positive values refer to tension (red) and negative values to compression (blue). Search band for stress maximum by white dotted line. (b) Profile of the tensile strength of slab layers (green area) and the tensile criterion T_i (black line). The orange area indicates the critical range below 1.

4. Results

We illustrate the slab tensile criterion with an example. Then, we provide comparisons of the measures of snow instability with experimental field data. Finally, we relate these instability metrics to the data set including signs of instability.

4.1. Slab Tensile Criterion

At the onset of crack propagation, that is, when a failure in the weak layer occurred and a crack begins to separate the weak layer, the slab undergoes strong deformation. The resulting stress in each slab layer depends on the layers' mechanical properties and was obtained from the FE simulation (Figure 1a). For each layer we compare the modeled maximum tensile stress to its strength. This ratio provides us with a tensile failure criterion T_i, which can be graphed along with a profile of strength (Figure 1b). Values of criterion T_i below 1 indicate that the layer is not strong enough to support the tensile stress. In the example (Figure 1b), this is the case for the top three layers of the slab yielding a slab tensile criterion T = 0.48. Indeed, in this particular case, we had observed a slab fracture in the PST at a crack length of 23 cm.

4.2. Comparing Metrics to Stability Test Results

In Figure 2a we compare the failure initiation criterion S modeled from SMP signals to corresponding RB scores (N = 62). As scores 1 and 7 were rarely observed, we pooled them with scores 2 and 6, respectively. Generally, the failure initiation criterion S increased with increasing RB score (r_s = 0.49), which is reflected in increasing medians in Figure 2a. Using the analytical solution for the maximum shear stress (McClung & Schweizer, 1999) instead of the FE solution decreased the correlation to r_s = 0.44 likely because slab layering is not appropriately taken into account (Figure S2).

Introducing a commonly used stability classification into stable (RB $\geq$ 4; N = 37) and unstable (RB $<$ 4; N = 25) cases (Schweizer & Jamieson, 2003), we found a split at S = 118 with a classification tree. At RPC = 0.73 (POD = 0.70 and PON = 0.75) the value separated well between stable and unstable cases. Apparently, slabs thicker than about 50 cm only rarely yielded low values of the failure initiation criterion, that is, $S < 118$, for which a weak layer strength below about 80 kPa was required.

The slab thickness (r_p = 0.76, $p < 0.01$) and the weak layer shear strength (r_p = 0.68, $p < 0.01$) were the main drivers of failure initiation and slightly more relevant than the FE-derived effective elastic modulus of the slab (r_p = 0.51, $p < 0.01$) assuming a linear correlation. As snowpack properties do not generally relate linearly to the metrics, we verified that applying a log transformation to nonlinear parameter relations does not change the above findings.

We compared the modeled critical crack length with the critical crack length observed in field experiments (PST data; Figure 2b, inset). Modeled and measured values of the critical crack length were clearly correlated (r_p = 0.84, $p < 0.01$). The model's potential for predictions is reflected in the coefficient of determination R^2 = 0.64. Indeed, the model could predict the critical crack length with a RMSE of 8 cm. As the mean

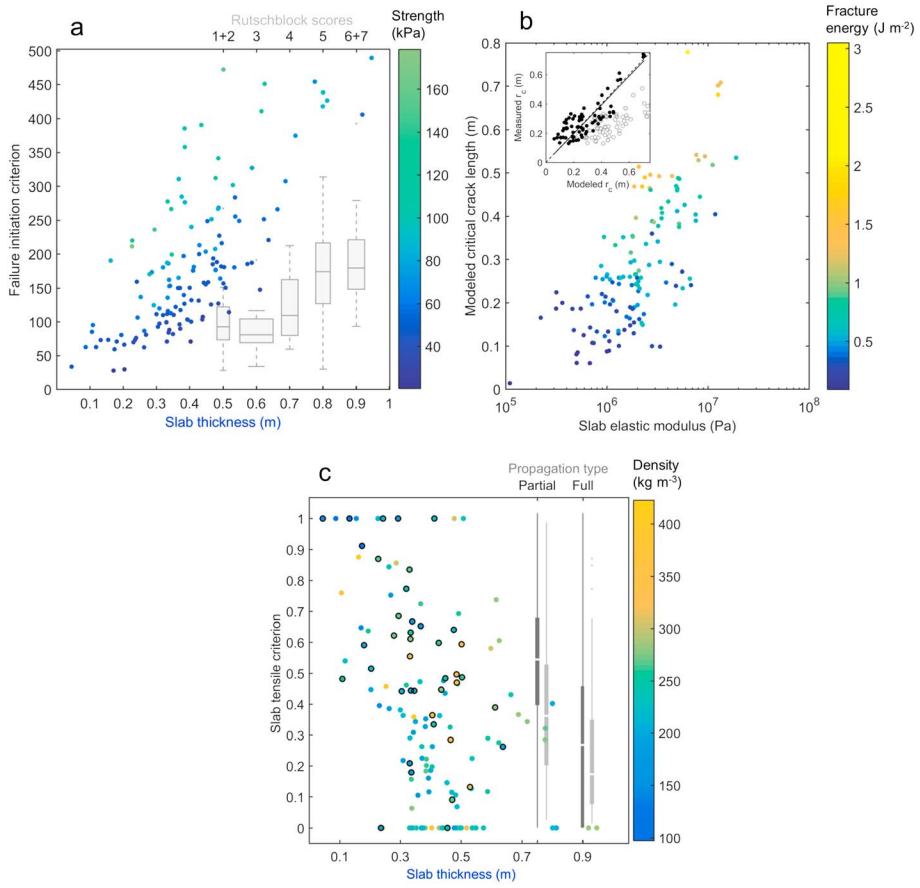

Figure 2. Metrics of instability compared to field tests (RB and PST data set, $N = 139$). (a) Failure initiation criterion versus slab thickness (colored dots; colors according to weak layer strength). Additionally, the failure initiation criterion is compared with Rutschblock scores from the RB data set ($N = 62$). The gray boxes span the interquartile range from first to third quartiles with a horizontal line showing the median. Whiskers extend to the most extreme data points not considered outliers (gray dots) within 1.5 times the interquartile range above the third quartile and below the first quartile. (b) Modeled critical crack length versus slab elastic modulus (colored dots; colors according to specific fracture energy of the weak layer). The inset compares modeled and measured critical crack lengths r_c from the PST data set ($N = 77$) considering slab layering with an effective slab modulus (solid black circles) or neglecting layering (open gray circles). (c) Slab tensile criterion versus slab thickness (colored dots; colors according to slab density). Black circles around the colored dots indicate *partial* propagation observed in field stability tests. The boxes show the FE based slab tensile criterion (dark gray boxes) and an analytical solution (light gray boxes) with observed *partial* and *full* propagation in field stability tests. Boxplots as described above in Figure 2a. RB = Rutschblock; PST = propagation saw test; FE = finite element.

absolute error was similar (7 cm), we conclude that with the current data the model performed robustly, that is, did not produce large outliers. Using a modulus based on the average slab density instead of an effective slab modulus was not a good approximation ($R^2 = -2.5$, RMSE = 23 cm; open gray circles).

The slab elastic modulus ($r_p = 0.71$, $p < 0.01$) and the weak layer specific fracture energy ($r_p = 0.83$, $p < 0.01$) both had a strong influence on the modeled critical crack length (Figure 2b). Critical crack lengths longer than 30 cm were only obtained if either the slab was rather stiff with slab elastic moduli larger than about 7 MPa or the weak layer was rather strong, which translates in our data set to specific fracture energies exceeding about 0.5 J/m^2.

To verify the slab tensile criterion T, we compared the modeled values with observations of the PST arrest type (Gauthier & Jamieson, 2008) and the RB release type (Schweizer et al., 2008). We formed two classes by assigning *whole block* release types in RB tests and *propagation to the end* in PSTs to the class *full* propagation ($N = 97$) and all other test results to the class *partial* propagation ($N = 42$).

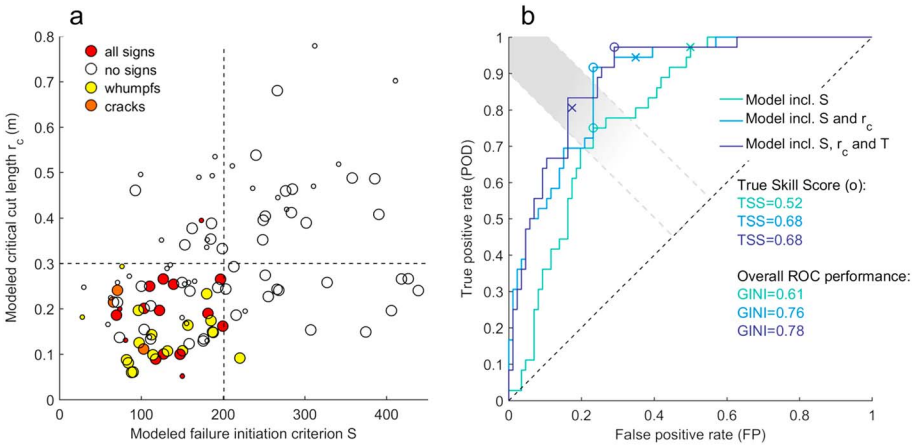

Figure 3. Validation and performance of metrics of instability. (a) Modeled failure initiation criterion, critical crack length, and slab tensile criterion (circle size: small circles indicating $T \geq 0.45$, large circles $T < 0.45$) are compared to independent observations of signs of instability ($N = 122$). Open circles indicate absence and colored circles indicate reported type of signs of instability. Dashed lines represent possible splits for S and r_c. (b) Presence of signs of instability in ROC space as predicted by models of increasing complexity (colors). Scores were calculated from logistic regressions including the failure initiation criterion (S) only, the latter plus the critical crack length (r_c), and the latter plus the slab tensile criterion (T). Circles indicate the highest value of TSS. The Gini coefficient summarizes model performance. Results with visually determined thresholds $S < 200$, $r_c < 30$ cm, and $T < 0.45$ are denoted by crosses. Shading illustrates region of balanced predictions.

The dark gray boxes in Figure 2c show that for lower values of the slab tensile criterion *T full* propagation was more likely observed ($r_s = -0.39$, $p < 0.01$). Indeed, a classification tree yielded a significant split for $T < 0.45$ between the groups of *partial* and *full* propagation. With an analytical approach based on the ratio of average slab tensile strength and slab tensile stress (Gaume et al., 2015) the correlation dropped to $r_s = -0.25$ (light gray boxes in Figure 2c). Hence, accounting for slab layering, the slab tensile criterion identified most cases showing *partial* propagation (POD = 0.64) and still kept them fairly separated from *full* propagation (PON = 0.75).

Whereas slab density was not related to the slab tensile criterion ($r_s = -0.07$, $p = 0.44$), the slab tensile criterion decreased with increasing slab thickness ($r_s = -0.44$, $p < 0.01$). Indeed, using the splitting value of $T < 0.45$, the slab tensile criterion suggests that slabs thicker than about 30 cm were required to support propagation.

4.3. A Model Framework of Snow Instability

Combining the modeled criteria, we aim at describing snow instability. For validation we used the data set with signs of instability ($N = 122$). We grouped observations of signs of instability into three groups: *whumpfs*, shooting cracks with or without whumpfs (*cracks*), and *all* (three) *signs* (Reuter, Schweizer, & van Herwijnen, 2015).

Apparently, signs of instability were concentrated in the lower left of the graph (Figure 3a). Indeed, visually selected thresholds of $S < 200$ and $r_c < 30$ cm separated 97% of the cases when signs of instability were present but observed stable cases, that is, cases when signs were absent, are poorly classified (RPC = 80%, POD = 0.94; PON = 0.65).

The lower left quadrant in Figure 3a holds 62 cases including 30 we missed to classify correctly, because signs of instability were absent. Among those false alarms were 16 cases that showed partial rather than *full* propagation in stability tests.

We showed above that the slab tensile criterion can discriminate between *full* and *partial* propagation. Introducing the slab tensile criterion with the before determined threshold of $T < 0.45$ led to the discrimination depicted by the small circles in Figure 3a. Consequently, the number of misclassified stable observations, that is, false alarms, fell by half, yielding a FP of 0.17 instead of 0.35. This, however, reduced the number of correctly identified unstable cases; the POD decreased from 0.94 to 0.81. Still, we obtained a more balanced classification with a slightly higher unweighted average classification accuracy RPC = 82%.

As the validation showed that thresholds can identify the unstable cases, that is, when signs of instability were reported (Figure 3a), we study the prediction performance of our three snow instability criteria for unstable cases. To do so, Figure 3b relates the false positive rate (FP) and the true positive rate (POD). Interpreting avalanche release as a sequence of fracture mechanical processes, we start with the failure initiation criterion and a threshold $S < 200$. Then we include the critical crack length with a threshold $r_c < 30$ cm and finally add the slab tensile criterion with a threshold $T < 0.45$. With increasing model complexity the overall prediction performance increased (crosses in Figure 3b), even though the models did not approach the upper left corner where a perfect model would sit. The model considering initiation only classified almost all unstable cases correctly (POD = 0.97), yet at a high cost (FP = 0.5). By including crack propagation and tensile support, we reduced the number of false stable predictions, that is, the costs, but at the same time the models become more liberal, meaning the false stable predictions increased.

We calculated ROC curves based on scores obtained from multiple logistic regressions. For one single classifier, such as the failure initiation criterion, points on the ROC curve represent different thresholds. Apparently, the visual threshold of $S < 200$ was a conservative choice, as it sits on the top center of the ROC curve and the highest true skill statistic (TSS = 0.52, open green circle in Figure 3b) is substantially to the lower left. This point corresponds to a threshold of $S < 152$, which is identical to the first split value found with a classification tree.

Including both, the critical crack length and the slab tensile criterion, increased the prediction performance. In particular, the area under the ROC curve extended further to the upper left that is reflected in Gini coefficients increasing from 0.61 to 0.76 and to 0.78. Although in some areas along the ROC curve the two-step model performed equally well as the three-step model, the three-step model performed slightly better in areas where stable and unstable cases are equally well predicted (see shading in Figure 3b)—hence yielding a more balanced prediction. This suggests that in our data set snow instability is best described by a combination of failure initiation, crack propagation, and slab tensile support.

5. Discussion

Whereas previous approaches usually relied on a single metric or snow property, we combined the three criteria for failure initiation, crack propagation, and tensile support to describe snow instability; our approach is guided by the fracture processes involved in slab avalanche release. The classical skier stability index combines weak layer strength, slab density, and slab thickness, which can be obtained with field measurements requiring detailed snow pit observations. The index agreed rather well with observed avalanche activity (Jamieson & Johnston, 1998), given that it only considers failure initiation. Bellaire et al. (2009) related SMP-derived strength and slab hardness separately to stability test results and concluded that snowpack stability is not simply a question of minimal strength but a complex interaction between weak layer and slab properties. Approximating the influence of the slab by considering slab density Pielmeier and Marshall (2009) improved the agreement of weak layer strength with field observations of snow instability. The first attempt to derive a snow instability metric from SMP measurements combining weak layer and slab properties was presented by Schweizer and Reuter (2015), but the relation to slope stability was poor, possibly because their approach did not consider crack propagation.

Previously presented new metrics describing both, failure initiation and crack propagation, helped to reproduce field observations, such as stability test results or signs of instability (Gaume & Reuter, 2017; Reuter, Schweizer, & van Herwijnen, 2015). By introducing the criterion for tensile support we improved the performance in describing snow instability. For this last analysis, data were already available from previous studies, which we updated with data we had collected meanwhile.

Due to this larger data set, recent developments in mechanical modeling (van Herwijnen et al., 2016) and an improved calibration of SMP-derived snow density, the thresholds of the instability criteria reported by Reuter, Schweizer, and van Herwijnen (2015), $S < 234$, $r_c < 41$ cm, slightly changed. Their findings on the role of failure initiation and crack propagation in slab avalanche release were confirmed despite some model adaptations. Moreover, including in the avalanche release processes the tensile support of the slab during crack propagation we are now able to account for a lack of slab strength—an effect that causes early slab fractures.

We used SMP penetration resistance data to derive mechanical properties and eventually calculate snow instability metrics. Hence, most properties, for example, the elastic modulus, are not directly measured but derived based on assumptions about how the tip of the SMP interacts with the snow. Presently, no comprehensive model exists that describes this interaction with the complex microstructure of snow. Nonetheless, comparisons with independent measurements, such as μCT imaging and FE modeling, revealed that the SMP-derived micromechanical properties are related to the macroscopic snow mechanical properties (Reuter et al., 2018). Still, to increase accuracy and robustness, we could average several SMP signals from the same location.

The presented framework offers a way to assess snow instability at a location in the field. However, we are lacking a reference for snow instability which in part is a consequence of slope scale effects (Gaume et al., 2014) and spatial variations (Reuter, Richter, & Schweizer, 2016). Moreover, to validate snow instability model results subjective field observations are typically only partly suitable. Hence, although we suggest objective metrics to determine snow instability, we are bound to use subjective observations for validation.

Addressing the present lack of an unambiguous terminology or a physical formulation of snow instability, this framework is a suggested concept of how to describe snow instability. As the framework is basically applicable to any kind of snow mechanical data, we suggest taking further steps toward the use with, for instance, snow cover modeling data and further assess its applicability and potential as part of an operational model.

Our performance parameters, for example, the RCP = 82% for our simple three-step model with visual thresholds, are not cross validated, since we are not presenting an operational model but are primarily showing how model performance improves with increasing complexity.

6. Conclusions

For assessing snow instability we suggest a framework including a combination of failure initiation, crack propagation, and slab tensile support. We verified the three metrics, which were determined from SMP-derived mechanical properties and FE modeling, with independent field observations.

Relating snowpack properties to the metrics revealed drivers of the fracture processes. For failure initiation those are slab thickness and weak layer strength; for crack propagation they are elastic modulus and weak layer fracture energy together with slab thickness and tensile strength for tensile support. All three criteria showed better performance if slab layering was taken into account, which often required FE modeling.

Finally, we assessed the predictive power of our three metrics in comparison with observed signs of instability. Model performance increased with increasing model complexity. The most balanced prediction is obtained if failure initiation, crack propagation, and slab tensile support are considered—suggesting that snow instability can be described by these three processes.

Future development should further assess applicability and potential for operational use.

Acknowledgments
B. R. has been funded by the Swiss National Science Foundation (P2EZP2_168896). Data are available at www.envidat.ch, https://doi.org/10.16904/envidat.40.

References

Bartelt, P., & Lehning, M. (2002). A physical SNOWPACK model for the Swiss avalanche warning; Part I: Numerical model. *Cold Regions Science and Technology*, *35*(3), 123–145. https://doi.org/10.1016/S0165-232X(02)00074-5

Bellaire, S., & Jamieson, B. (2013). Forecasting the formation of critical snow layers using a coupled snow cover and weather model. *Cold Regions Science and Technology*, *94*, 37–44. https://doi.org/10.1016/j.coldregions.2013.06.007

Bellaire, S., Pielmeier, C., Schneebeli, M., & Schweizer, J. (2009). Stability algorithm for snow micro-penetrometer measurements. *Journal of Glaciology*, *55*(193), 805–813. https://doi.org/10.3189/002214309790152582

Birkeland, K. W. (2001). Spatial patterns of snow stability throughout a small mountain range. *Journal of Glaciology*, *47*(157), 176–186. https://doi.org/10.3189/172756501781832250

Brun, E., David, P., Sudul, M., & Brunot, G. (1992). A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting. *Journal of Glaciology*, *38*(128), 13–22. https://doi.org/10.1017/S0022143000009552

Canadian Avalanche Association (2007). Revelstoke BC, Canada: Canadian Avalanche Association (CAA).

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Föhn, P. M. B. (1987). The Rutschblock as a practical tool for slope stability evaluation. In B. Salm & H. Gubler (Eds.), *Symposium at Davos 1986 - Avalanche Formation, Movement and Effects* (pp. 223–228), IAHS Publ., 162. Wallingford, Oxfordshire, UK: International Association of Hydrological Sciences.

Gaume, J., & Reuter, B. (2017). Assessing snow instability in skier-triggered snow slab avalanches by combining failure initiation and crack propagation. *Cold Regions Science and Technology*, *144*, 6–15. https://doi.org/10.1016/j.coldregions.2017.05.011

Gaume, J., Schweizer, J., van Herwijnen, A., Chambon, G., Reuter, B., Eckert, N., & Naaim, M. (2014). Evaluation of slope stability with respect to snowpack spatial variability. *Journal of Geophysical Research: Atmospheres*, *119*, 1783–1799. https://doi.org/10.1002/2014JF003193

Gaume, J., van Herwijnen, A., Chambon, G., Birkeland, K. W., & Schweizer, J. (2015). Modeling of crack propagation in weak snowpack layers using the discrete element method. *The Cryosphere*, *9*(5), 1915–1932. https://doi.org/10.5194/tc-9-1915-2015

Gauthier, D., & Jamieson, B. (2008). Evaluation of a prototype field test for fracture and failure propagation propensity in weak snowpack layers. *Cold Regions Science and Technology*, *51*(2–3), 87–97. https://doi.org/10.1016/j.coldregions.2007.04.005

Heierli, J., Gumbsch, P., & Zaiser, M. (2008). Anticrack nucleation as triggering mechanism for snow slab avalanches. *Science*, *321*(5886), 240–243. https://doi.org/10.1126/science.1153948

Jamieson, J. B., Haegeli, P., & Schweizer, J. (2009). Field observations for estimating the local avalanche danger in the Columbia Mountains of Canada. *Cold Regions Science and Technology*, *58*(1–2), 84–91. https://doi.org/10.1016/j.coldregions.2009.03.005

Jamieson, J. B., & Johnston, C. D. (1990). In-situ tensile tests of snowpack layers. *Journal of Glaciology*, *36*(122), 102–106. https://doi.org/10.1017/S002214300000561X

Jamieson, J. B., & Johnston, C. D. (1998). Refinements to the stability index for skier-triggered dry slab avalanches. *Annals of Glaciology*, *26*, 296–302. https://doi.org/10.1017/S0260305500014993

Johnson, J. B., & Schneebeli, M. (1999). Characterizing the microstructural and micromechanical properties of snow. *Cold Regions Science and Technology*, *30*(1–3), 91–100. https://doi.org/10.1016/S0165-232X(99)00013-0

Kronholm, K., & Schweizer, J. (2003). Snow stability variation on small slopes. *Cold Regions Science and Technology*, *37*(3), 453–465. https://doi.org/10.1016/S0165-232X(03)00084-3

Löwe, H., & van Herwijnen, A. (2012). A Poisson shot noise model for micro-penetration of snow. *Cold Regions Science and Technology*, *70*, 62–70. https://doi.org/10.1016/j.coldregions.2011.09.001

McClung, D. M., & Schweizer, J. (1999). Skier triggering, snow temperatures and the stability index for dry slab avalanche initiation. *Journal of Glaciology*, *45*(150), 190–200. https://doi.org/10.1017/S0022143000001696

Monti, F., Gaume, J., van Herwijnen, A., & Schweizer, J. (2016). Snow instability evaluation: Calculating the skier-induced stress in a multi-layered snowpack. *Natural Hazards and Earth System Sciences*, *16*(3), 775–788. https://doi.org/10.5194/nhess-16-775-2016

Pielmeier, C., & Marshall, H.-P. (2009). Rutschblock-scale snowpack stability derived from multiple quality-controlled SnowMicroPen measurements. *Cold Regions Science and Technology*, *59*(2–3), 178–184. https://doi.org/10.1016/j.coldregions.2009.06.005

Proksch, M., Löwe, H., & Schneebeli, M. (2015). Density, specific surface area and correlation length of snow measured by high-resolution penetrometry. *Journal of Geophysical Research: Earth Surface*, *120*, 346–362. https://doi.org/10.1002/2014JF003266

Quéno, L., Vionnet, V., Dombrowski-Etchevers, I., Lafaysse, M., Dumont, M., & Karbou, F. (2016). Snowpack modelling in the Pyrenees driven by kilometric-resolution meteorological forecasts. *The Cryosphere*, *10*(4), 1571–1589. https://doi.org/10.5194/tc-10-1571-2016

Reuter, B., van Herwijnen, A., Bellaire, S., & Schweizer, J. (2018). Data set on snow instability. WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland. https://doi.org./10.16904/envidat.40

Reuter, B., Richter, B., & Schweizer, J. (2016). Snow instability patterns at the scale of a small basin. *Journal of Geophysical Research: Earth Surface*, *121*, 257–282. https://doi.org/10.1002/2015JF003700

Reuter, B., Schweizer, J., & van Herwijnen, A. (2015). A process-based approach to estimate point snow instability. *The Cryosphere*, *9*(3), 837–847. https://doi.org/10.5194/tc-9-837-2015

Reuter, B., van Herwijnen, A., & Schweizer, J. (2016). Observer independent measures of snow instability. In E. Greene (Ed.), *Proceedings ISSW 2016. International Snow Science Workshop, Breckenridge CO, U.S.A., 3–7 October 2016* (pp. 390–396).

Reuter, B., van Herwijnen, A., Veitinger, J., & Schweizer, J. (2015). Relating simple drivers to snow instability. *Cold Regions Science and Technology*, *120*, 168–178. https://doi.org/10.1016/j.coldregions.2015.06.016

Schneebeli, M., & Johnson, J. B. (1998). A constant-speed penetrometer for high-resolution snow stratigraphy. *Annals of Glaciology*, *26*, 107–111. https://doi.org/10.1017/S0260305500014658

Schweizer, J., & Jamieson, J. B. (2003). Snowpack properties for snow profile analysis. *Cold Regions Science and Technology*, *37*(3), 233–241. https://doi.org/10.1016/S0165-232X(03)00067-3

Schweizer, J., & Jamieson, J. B. (2010). Snowpack tests for assessing snow-slope instability. *Annals of Glaciology*, *51*(54), 187–194. https://doi.org/10.3189/172756410791386652

Schweizer, J., Jamieson, J. B., & Schneebeli, M. (2003). Snow avalanche formation. *Reviews of Geophysics*, *41*(4), 1016. https://doi.org/10.1029/2002RG000123

Schweizer, J., McCammon, I., & Jamieson, J. B. (2008). Snowpack observations and fracture concepts for skier-triggering of dry-snow slab avalanches. *Cold Regions Science and Technology*, *51*(2–3), 112–121. https://doi.org/10.1016/j.coldregions.2007.04.019

Schweizer, J., & Reuter, B. (2015). A new index combining weak layer and slab properties for snow instability prediction. *Natural Hazards and Earth System Sciences*, *15*(1), 109–118. https://doi.org/10.5194/nhess-15-109-2015

Schweizer, J., Reuter, B., van Herwijnen, A., Jamieson, J. B., & Gauthier, D. (2014). On how the tensile strength of the slab affects crack propagation propensity. In Haegeli, P. (Ed.), *Proceedings ISSW 2014. International Snow Science Workshop, Banff, Alberta, Canada, 29 September – 3 October 2014* (pp. 164–168).

Schweizer, J., Reuter, B., van Herwijnen, A., Richter, B., & Gaume, J. (2016). Temporal evolution of crack propagation propensity in snow in relation to slab and weak layer properties. *The Cryosphere*, *10*(6), 2637–2653. https://doi.org/10.5194/tc-10-2637-2016

Techel, F., & Schweizer, J. (2017). On using local avalanche danger level estimates for regional forecast verification. *Cold Regions Science and Technology*, *144*, 52–62. https://doi.org/10.1016/j.coldregions.2017.07.012

van Herwijnen, A., Gaume, J., Bair, E. H., Reuter, B., Birkeland, K. W., & Schweizer, J. (2016). Estimating the effective elastic modulus and specific fracture energy of snowpack layers from field experiments. *Journal of Glaciology*, *62*(236), 997–1007. https://doi.org/10.1017/jog.2016.90

van Herwijnen, A., & Jamieson, J. B. (2007). Snowpack properties associated with fracture initiation and propagation resulting in skier-triggered dry snow slab avalanches. *Cold Regions Science and Technology*, *50*(1–3), 13–22. https://doi.org/10.1016/j.coldregions.2007.02.004

Vernay, M., Lafaysse, M., Mérindol, L., Giraud, G., & Morin, S. (2015). Ensemble forecasting of snowpack conditions and avalanche hazard. *Cold Regions Science and Technology*, *120*, 251–262. https://doi.org/10.1016/j.coldregions.2015.04.01

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., et al. (2012). The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2. *Geoscientific Model Development*, *5*(3), 773–791. https://doi.org/10.5194/gmd-5-773-2012

Vionnet, V., Dombrowski-Etchevers, I., Lafaysse, M., Quéno, L., Seity, Y., & Bazile, E. (2016). Numerical weather forecasts at kilometer scale in the French Alps: Evaluation and application for snowpack modeling. *Journal of Hydrometeorology*, *17*(10), 2591–2614. https://doi.org/10.1175/JHM-D-15-0241.1

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (3rd ed., Vol. 100). San Diego, CA: Academic Press.