Hardness estimation and weak layer detection in simulated snow stratigraphy

CrossMark

Fabiano Monti *, Jürg Schweizer, Charles Fierz

WSL Institute for Snow and Avalanche Research SLF, Davos Dorf, Switzerland

ARTICLE INFO

ABSTRACT

Numerical modeling of snow cover stratigraphy with, for example, the 1-D snow cover model SNOWPACK has the potential to increase the spatial and temporal resolutions of snow stratigraphy information — data very much needed for avalanche forecasting. One of the key properties for interpreting snow stratigraphy in regard to stability is snow hardness. In manually observed snow profiles, differences in snow hardness between layers were found to be indicative of instability. We improved the hardness parameterization implemented in the snow cover model SNOWPACK. Hardness is estimated from the simulated snow density and grain type. Density thresholds for primary grain types and all hardness steps were calculated using ordinal logistic regression (on a data set of 14,522 manually observed layers). We thus implemented snow hardness as a discrete parameter in SNOWPACK. The new hardness parameterization observed agreed well with the simulated snow hardness. The structural stability index (SSI), and the threshold sum approach (TSA) were then used to detect potential weak layers in the simulated stratigraphy. We evaluated whether failure layers detected with compression tests (CT) in manually observed snow profiles corresponded to the potential weak layers found by either the SSI or TSA in the simulated stratigraphy. CT failure layers corresponded in about half of the cases to the potential weak layers detected with either the SSI or the TSA in the simulated stratigraphy. The agreement improved if only sudden collapse fractures were considered. These findings suggest that stability information can be derived from simulated snow stratigraphy in particular if the method for detecting weak layers is further improved.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

A prerequisite for dry-snow slab avalanches is a snowpack weakness below one or more cohesive slab layers. The susceptibility of a weak layer to failure initiation and crack propagation determines snow instability (Schweizer et al., 2003). Identifying potential weak layers is therefore among the key tasks in snow stability assessment. Estimating snow stability is widely based on snow profile interpretation, which is a fairly subjective method, especially when stability tests are not available (Schweizer and Wiesinger, 2001). To quantify profile interpretation Schweizer and Jamieson (2007) proposed a threshold sum approach (TSA) that evaluates six structural variables, and is similar to the so-called lemons introduced by McCammon and Schweizer (2002). While the number of variables in the critical range correlated with instability for known failure layers, the method provided poor results if used for detecting potential weak layers (Schweizer and Jamieson, 2007); it overestimated instability (Winkler and Schweizer, 2009). A widely used stability test is the compression test (CT) (CAA, 2007), which

similarly to the TSA well detects weak layers (high sensitivity), but has a high false alarm ratio (low specificity) (Schweizer and Jamieson, 2010).

As field observations are time consuming and sometimes not feasible due to avalanche danger, assessing snow stability from simulated snow stratigraphy would help to increase snow cover information in space and time. Whereas manually observed snow profiles are supplemented with a snow stability test to facilitate profile interpretation, this option is not available when interpreting simulated snow stratigraphy. Therefore, mechanical or structural parameters are used to find potential weak layers and assess their strength (e.g. Durand et al., 1999; Lehning et al., 2004). Several studies related output from the 1-D numerical snow cover model SNOWPACK (Lehning et al., 2002a,b) to observed stability. Schweizer et al. (2006) proposed a structural stability index (SSI) which combines two structural instability parameters (difference in grain size and hardness between adjacent layers) with the classical skier stability index (SK38) introduced by Föhn (1987) and refined by Jamieson and Johnston (1998). The SSI proved to be superior to the SK38 in detecting failure layers (Schweizer et al., 2006). Recently, Monti et al. (2012) related modeled stability information derived with the TSA to regional avalanche danger.

While these studies demonstrated a link between modeled snowpack variables and observed signs of instability such as avalanche

* Corresponding author at: WSL Institute for Snow and Avalanche Research, Flüelastrasse 11, CH-7260 Davos Dorf, Switzerland. Tel.: +41 81 4170 252.
E-mail address: monti@slf.ch (F. Monti).

activity, snowpack stability estimates or avalanche danger, a profile-by-profile comparison between observed failure layers and simulated weak layers is still lacking.

The aim of this study is therefore to verify whether (1) observed failure layers can be detected in snow profiles simulated with the snow cover model SNOWPACK, and (2) if potential weak layers identified in simulated snow profiles correspond to observed unstable layers. We compared failure layers found with compression tests to potential weak layers identified in the simulated snow stratigraphy with either the structural stability index (SSI) or the threshold sum approach (TSA), and vice versa. Since both the SSI and the TSA strongly depend on snow hardness, we first had to refine the parameterization of the hand hardness index within SNOWPACK.

2. Data

2.1. Hand hardness calibration

For the hand hardness parameterization we used two snow profile data sets collected by snow observers of the WSL Institute for Snow and Avalanche Research SLF and the ARPAV Avalanche Center of Arabba (north-eastern Italian Alps). The Swiss data set consisted of 2349 observed snow layers from 312 snow profiles recorded in the area surrounding Davos (from 1560 m a.s.l. to 2810 m a.s.l.). The Italian data set consisted of 12,173 snow layers from 1125 profiles collected in the whole Veneto Region (from 1170 m a.s.l. to 2940 m a.s.l.) during the winters 1999–2000 to 2011–2012. For both data sets hand hardness was related to density depending on the grain type. In order to assess the suitability of the new hardness parameterization, the Swiss hardness data were compared to modeled hardness (12,437 layers) from 143 profiles (simulated for the locations of Weissfluhjoch and Steintälli) covering about the same area and time as the observed ones. These data are mostly similar to the data used by Schweizer et al. (2006) and their characteristics can be found therein. Our analysis was confined to only dry snow layers.

2.2. Weak layer detection

Meteorological input data measured at the two Automated Weather Stations (AWS), Weissfluhjoch (2540 m a.s.l.) and Steintälli (2440 m a.s.l.), above Davos, Switzerland, were used to simulate snow stratigraphy with the 1-D model SNOWPACK. We compared these simulations with 83 manually observed snow profiles each included at least one CT. In total 180 failure layers were found using the CT; for only 129 the fracture character (van Herwijnen and Jamieson, 2007) was indicated. The field data were collected in the flat study plots surrounding the two AWS during the winters 1999–2000 to 2011–2012. In total 1790 manually observed snow layers could be compared to 7926 simulated layers.

3. Methods

In the following, we will first describe how we developed the new hardness parameterization. Then, we will discuss the methods for failure layer detection in manually observed and simulated snow stratigraphy. We compared the CT, SSI and TSA within and between manually observed and simulated profiles (Fig. 1). One can ask whether an observed failure layer is represented in the simulated snow stratigraphy, or vice versa, whether for a potential weakness in a simulated profile a corresponding failure layer was observed. Both directions of comparison were performed for assessing the value of the simulated snow stratigraphy for evaluating snow instability.

3.1. Hand hardness calibration

Snow hardness is credited as one of the most important parameters to assess snow stability (Pielmeier and Schneebeli, 2003b). Several types of hardness tests were developed in the last 80 years. Despite its subjectivity (Pielmeier and Schneebeli, 2003a), the hand hardness test is still the most widely used. Hand hardness is estimated by gently pushing the fist, four fingers, one finger, a pencil or a knife into to the
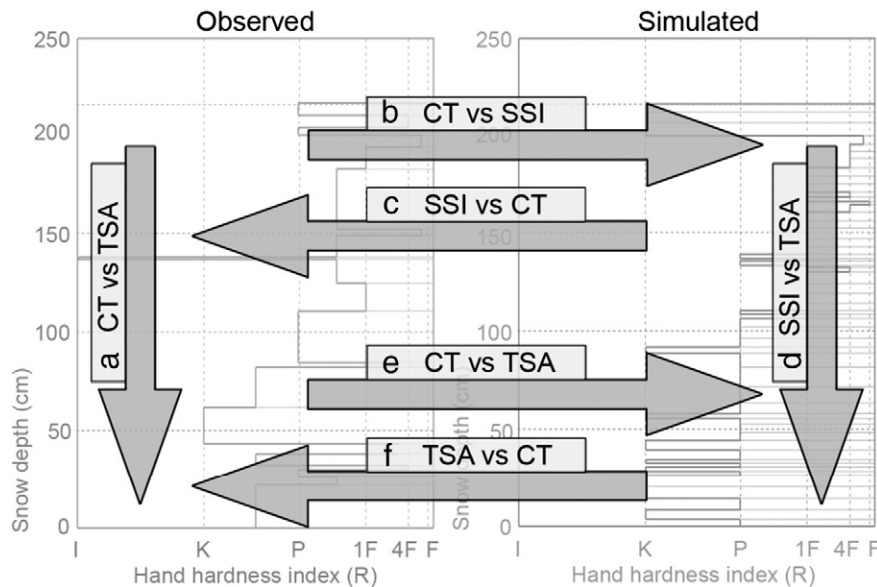

Fig. 1. Overview of the comparisons performed within and between observed and simulated profiles. Arrows are labeled a) to f): a) CT failure layers versus the TSA for observed profiles; b) CT failure layer versus a potential weak layer identified by the SSI at the corresponding depth in the simulated profiles; c) potential weak layer identified by the SSI versus CT failure layer at the corresponding depth in the observed profiles; d) potential weak layer identified by the SSI versus the TSA in the simulated profiles; e) CT failure layers versus a potential weak layer identified by the TSA at the corresponding depth in the simulated profiles; f) potential weak layer identified by the TSA versus CT failure layers at the corresponding depth in the observed profiles.

snow, thereby not exceeding a penetration force of 10-15 N (Fierz et al., 2009).

The snow hardness parameterization originally implemented in SNOWPACK was based on the relation between snow density and hand hardness index depending on grain type as suggested by Geldsetzer and Jamieson (2001). Recently, a new snow settlement parameterization was introduced into SNOWPACK which affects snow density simulation and thus the snow hardness estimation. Schweizer et al. (2006) had refined the original parameterization of snow hardness performing a multivariate statistical regression analysis for the snow hardness index for each grain type using snow density and grain type as independent variables. However, the differences between simulated and observed snow densities strongly affected the calculation of the hardness index. To address this issue Schweizer et al. (2006) scaled simulated to observed snow densities to improve the agreement. This method has among others two major disadvantages: i) since it is derived with linear regression the simulated hand hardness becomes a continuous variable, while for the observer it is a discrete parameter (Fierz et al., 2009); ii) adjusting simulated densities to measured ones implies that linear regression results are applicable only for that particular version of the model; whenever an improvement affecting snow density or grain type assignment is implemented, a re-calibration is required.

To preserve the discrete character of the hand hardness index, we performed multiple ordinal logistic regressions (McCullagh and Nelder, 1990) maintaining snow density and grain type as independent variables. We considered the following seven grain types for the analysis: precipitation particles (PP), decomposing and fragmented precipitation particles (DF), rounded grains (RG), faceted crystals (FC), depth hoar (DH), melt forms (MF) and rounding faceted particles (FCxr). The Swiss dataset included half steps of the hand hardness index (e.g. 1 to 2). To combine the regression results of the two datasets, the Swiss data had to be assigned to classes without half steps (e.g. 1 or 2). The half step classes were allocated to full steps considering the overlap of the density distribution (e.g. 1 to 2) with the density distributions of the main classes (e.g. 1 and 2). The combined regression results were then applied to the simulated layers. Simulated hand hardness for each snow layer was obtained from the mean between the hand hardness calculated separately for the primary and the secondary grain type. Consequently, half steps of hand hardness index can be obtained, which we consider as useful in view of snow stability evaluation. For example, for two layers of similar density, one consisting of faceted crystals and rounded grains, the other of faceted crystals only, our approach will result in different values of the hand hardness index.

Simulating snow hardness as a discrete (ordinal) value makes it less sensitive to differences between measured and simulated snow density. For this reason we did not scale the simulated densities to the measured ones; consequently snow model improvements are possible without the need to re-calibrate the hardness index.

The non-parametric Mann–Whitney U-Test was used to contrast variables (e.g. density and hand hardness) from simulated to observed data sets; differences were judged to be statistically significant for $p < 0.05$.

3.2. Potential weak layers in manual snow profiles

To detect potential weak layers in manually observed snow profiles the CT was used. The CT has been validated for tests performed on slopes (Jamieson, 1999). To our knowledge, no validation results exist for CT's performed in flat field study plots. Nevertheless, we assumed that potential weak layers can as well be detected with the CT in level terrain, which is in accordance with common practice (e.g. CAA, 2007). Compression test results can be further classified based on the fracture character as suggested by van Herwijnen and Jamieson (2007): progressive compression (PC), resistant planar (RP), sudden planar (SP), sudden collapse (SC) and non-planar break (B). However, in our data set (going back to the winter 1999–2000) the fracture character was not always recorded.

The threshold sum approach (Schweizer and Jamieson, 2007) was used to verify if a failure layer detected with the CT could have been found using only stratigraphy information (a in Fig. 1). The TSA is generally calculated for all interfaces within the profile. Six snow stratigraphy variables were found to be related with snow stability (Schweizer and Jamieson, 2007): 1) difference in grain size; 2) failure layer grain size; 3) difference in hardness; 4) failure layer hardness; 5) failure layer grain type; 6) failure layer depth. If five or six of these variables were within their respective critical range (Table 1), Schweizer and Jamieson (2007) considered the corresponding interface as potentially unstable. Since we focus on potential weak layers (not interfaces) the critical variables were assigned to one of the adjacent layers as described in Monti et al. (2012). The threshold for failure layer depth was adapted to ease comparison with the CT results; failures in the topmost 18 cm of the snowpack were not excluded (Table 1).

3.3. Potential weak layers in simulated snow profiles

To detect potential weak layers in simulated snow stratigraphy we applied both the SSI (Schweizer et al., 2006) and the TSA since stability test information is obviously missing. The SSI is calculated for each layer boundary as:

$$SSI = SK38 + D \tag{1}$$

with

$$SK38 = \frac{\tau_{I,II}}{\tau_{xz} + \Delta\tau_{xz}} \tag{2}$$

where τ_I and τ_{II} are the shear strength for persistent and non-persistent grain types, respectively (Jamieson and Johnston, 1998), τ_{xz} is the shear stress due to the weight of the overlaying slab and $\Delta\tau_{xz}$ is the additional shear stress due to the skier (Föhn, 1987) that includes several refinements such as the effect of ski penetration (Jamieson and Johnston, 2001). D is defined as:

$$D = \begin{cases} 0 \text{ if } \Delta R \geq 1.5 \text{ and } \Delta E \geq 0.5 \text{ mm} \\ 1 \text{ if } \Delta R < 1.5 \text{ or } \Delta E < 0.5 \text{ mm} \\ 2 \text{ if } \Delta R < 1.5 \text{ and } \Delta E < 0.5 \text{ mm} \end{cases} \tag{3}$$

where ΔR is the absolute difference of the hand hardness index and ΔE is the absolute difference in grain size between two adjacent layers.

For each simulated snow profile, the interface with the minimum value of SSI was selected as the most probable potential weak interface. The softer of the two layers adjacent to the interface was considered as the potential weak layer and used for comparison with the manually observed profiles.

The threshold sum approach (TSA) was also applied to analyze the simulated profiles. The critical variables were assigned to the layers as described above (see Section 3.2) for the manually observed profiles. The thresholds proposed by Monti et al. (2012) (Table 1) for snow

Table 1
Critical ranges of variables for calculating the stratigraphical threshold sum approach (TSA). Thresholds for manually observed profiles proposed by Schweizer and Jamieson (2007) (Orig. obs.) and used in this work (observed) as well as the thresholds for simulated profiles proposed by Monti et al. (2012) (Orig. sim.) and used in this study (simulated) are given.

Variable or classifier	Threshold value			
	Orig. obs.	Observed	Orig. sim.	Simulated
Failure layer grain size	≥1.25 mm	≥1.25 mm	>0.6 mm	>0.6 mm
Difference in grain size	≥0.75 mm	≥0.75 mm	≥40%	≥40%
Difference in hardness	≥1.7	≥1.7	≥1	≥1.7
Failure layer hardness	≤1.3	≤1.3	≤2	≤1.3
Failure layer grain type	Persistent	Persistent	Persistent	Persistent
Failure layer depth	18...94 cm	≤100 cm	≤100 cm	≤100 cm

hardness were modified as a new hardness parameterization was introduced. Since snow hardness in SNOWPACK is now estimated as a discrete variable as observed in the field, thresholds were set equal to those used for manually observed profiles (Table 1).

3.4. SSI validation

First of all, we compared the characteristics of failure layers found by the CT in the manually observed profiles with the characteristics of the potential weak layers detected by the SSI in the simulated profiles. Then we checked, profile by profile, if for observed CT failure layers a corresponding potential weak layer was found by the SSI (b in Fig. 1). The probability of detection (POD) was used as a performance measure (Wilks, 1995). Lehning et al. (2001) proposed an objective method to compare simulated and observed snow profiles. This method can be adapted to relate the failure layers detected by the CT with the potential layers detected at about the same depth in the simulations, or vice versa (c in Fig. 1).

First of all, the height of the modeled interfaces z'^{mod}_i is stretched linearly in order to compensate for a possible difference in total snow height (obtaining z^{mod}_i):

$$z^{mod}_i = z'^{mod}_i \frac{z^{obs}_{nO}}{z'^{mod}_{nM}}.$$ (4)

Where z^{obs}_{nO} and z'^{mod}_{nM} are the total snow height in observed and simulated profiles, respectively. The index i of observed and modeled layer boundaries goes from 0 (at the bottom) to nO and nM (at the snow surface), respectively.

Due to the fact that in simulated profiles the number of layers is typically larger than in observed profiles, a height range around the CT failure layer is calculated; the corresponding simulated layer has to be found within this range. The tolerance of the height range is a function of the total snow depth and location within the snow cover (Lehning et al., 2001). The lower h^{mod}_{lower} and upper h^{mod}_{upper} heights of the range are defined as follows:

$$h^{mod}_{lower} = z^{obs}_i + Wt_i; h^{mod}_{upper} = z^{obs}_{i+1} + Wt_{i+1}$$ (5)

with

$$t_i = \frac{\left(z^{obs}_{nO} - z^{obs}_i\right) z^{obs}_i}{z^{obs}_{nO}}.$$ (6)

With the manual profiles as reference, Lehning et al. (2001) proposed $W = 0.2$ as suitable tolerance factor. Once the maximum range of tolerance around a CT failure layer is defined, the presence (within the tolerance range) of corresponding simulated layers detected as potentially unstable with the SSI is assessed. In order to have a similar range of tolerance for both manually observed and simulated profiles, the tolerance factor has to be larger for the simulated profiles than for the manually observed ones because layers in the simulated profiles are significantly thinner; we therefore used $W = 0.5$ for the simulated profiles.

To evaluate the stability information provided by the SSI and the TSA for the simulated profiles the two methods are compared (d in Fig. 1).

3.5. TSA validation

To validate the threshold sum approach, first, manually observed snow profiles were selected as the reference. Failure layers found with the CT in the manually observed profiles were used to verify if snow cover weaknesses were detected by the TSA at the corresponding depths in the simulated profiles (e in Fig. 1). Again, the above described comparison method proposed by Lehning et al. (2001) was applied.

Possible differences in total snow depth were taken into account following Eq. (4) and tolerance depth ranges were calculated according to Eqs. (5) and (6). Again, alternatively, the simulated profiles were considered as the reference. We hence checked whether for a potential weak layer identified by the TSA in the simulated profiles a corresponding CT failure layer can be found in the manually observed profile (f in Fig. 1). The POD was used to assess the performance.

4. Results

4.1. Hardness simulation

Verifying the agreement between simulated and observed snow densities is fundamental for subsequently analyzing the hand hardness index estimation. Modeled densities were generally higher than observed ones for all grain types except FC and MF (Fig. 2). Contrasting the measured to simulated densities showed that for all grain types the differences between the samples were statistically significant (U-Test, $p < 0.05$).

The simulated snow hardness was derived from the snow density thresholds, which were defined from the hand hardness index steps per grain type found with the multiple ordinary logistic regressions (Table 2). For example, Fig. 3 shows the resulting probability distributions for the grain type FC. For a given snow density, e.g. 300 kg/m³, the probabilities for the hand hardness steps 4F, 1F, and P or harder are 0.55, 0.27 and 0.18, respectively. Correspondingly, in Table 2, the lower and upper thresholds for hand hardness index 2 are 248 and 319 kg/m³. As illustrated in Fig. 3 the hardness variation for a given density is large. Overall, the probability of detection was 0.83, 0.73, 0.49, 0.53, 0.55, 0.28 and 0.47 for the grain types PP, DF, RG, FC, DH, MF, and FCxr, respectively. The median values of the simulated and observed snow hardness index for the grain types PP, DF and FC (Fig. 4) agreed. For the grain types RG, DH, MF and FCxr the median of the simulated hand hardness values was half a step higher than the median of the observed hand hardness. These differences are due to the aforementioned density difference between model and observations (Fig. 2). For all grain types, the distributions for observed and simulated hardness were found to be significantly different (p < 0.05).

4.2. SSI validation

A total of 83 snow profiles with a compression test were related to the corresponding simulated profiles; the observed profiles were taken as the reference. The relative distributions of grain types in manually observed and simulated layers were similar (Fig. 5) in spite of the large difference in terms of the number of layers (1790 manually
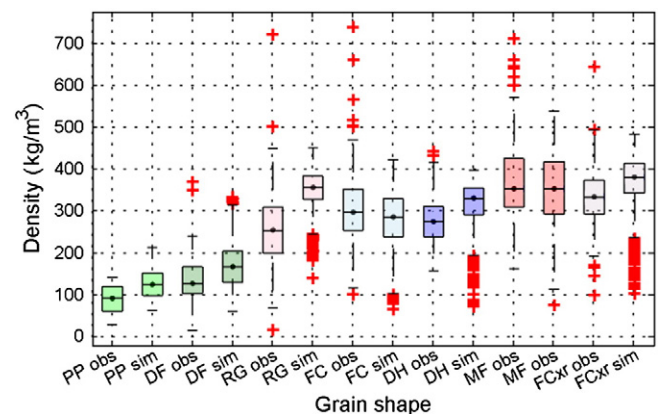

Fig. 2. Observed (obs) vs. modeled (sim) density distributions per grain type. Boxes span the interquartile range from 1st to 3rd quartile with line including a dot showing the median. Whiskers show the range of observed values that fall within ±1.5 times the interquartile range; crosses show outliers.

Table 2
Lower and upper threshold values of snow density (kg/m³) for each hand hardness index (R = 1 … 5) according to different grain types. The value 0 kg/m³ was set as arbitrary lower boundary for hand hardness index R = 1.

Hand hardness	(R)	1	2	3	4	5
PP	Lower	0	144	–	–	–
	Upper	143	–	–	–	–
DF	Lower	0	214	269	–	–
	Upper	213	268	–	–	–
RG	Lower	0	190	278	369	444
	Upper	189	277	368	443	–
FC	Lower	0	248	320	401	519
	Upper	247	319	400	518	–
DH	Lower	0	288	345	–	–
	Upper	287	344	–	–	–
MF	Lower	0	214	318	408	740
	Upper	213	317	407	739	–
FCxr	Lower	0	260	327	397	485
	Upper	259	326	396	484	–

Fig. 4. Observed (obs) vs. modeled (sim) hand hardness distributions per grain type. Same representation as in Fig. 2.

observed vs. 7926 simulated). In the simulated profiles, fewer SH layers were found. This finding might be due to the tendency of the model to rapidly transform SH layers, once buried into either layers of DH or FC. In Table 3 some layer characteristics (frequency of grain type, median grain size, median hand hardness index, median number of structural variables in the critical range) are summarized for observed and simulated layers. The agreement between the observed and simulated characteristics was in general satisfactory. The largest differences were found (a) for layers consisting of depth hoar crystals (DH): simulated grains were substantially smaller than observed ones, and (b) for layers consisting of melt forms (MF): simulated layers were harder than observed ones. The latter difference is an artifact within SNOWPACK since melt forms that have undergone one melt–freeze cycle are assigned a hand hardness index of 5 by default.

In Fig. 6 the frequency of grain types for all the 180 CT failure layers and the potential weak layers found by the SSI are shown. In the simulated profiles, faceted crystals and depth hoar were most frequently (in about 75% of the cases) found in potential weak layers. For the CT failure layers a greater variety of grain types was recorded. In the observed profiles, one fourth of the weak layers contained grain types related to recent snow (PP and DF). The absence of recent snow in potential weak layers detected by the SSI in simulated profiles is due to the fact that in SNOWPACK the grain size for PP is by default equal to 0.3 mm. Accordingly, DF grains are still small and the difference in grain size becomes often small. Consequently, the SSI rarely identified layers of DF's as weak layers due to the grain size threshold set for D (Eq. (3)). There were only two SH layers identified as potential weak
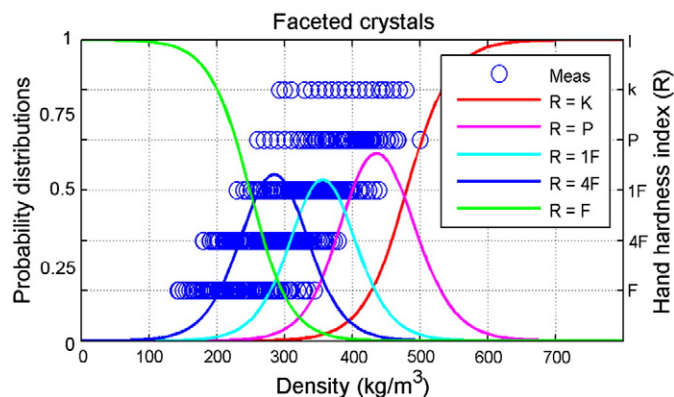
layers by the SSI; as previously mentioned this is related to the way snow metamorphism is implemented in SNOWPACK: after burial layers of SH are often quickly transformed in DH (or FC) layers. In general, the SSI (due to its definition) favors the detection of weaknesses related to persistent instabilities.

For each observed profile, the failure layers found with the CT were compared to the corresponding layers in the simulated profiles (b in Fig. 7). As only one potential weak layer can be detected with the SSI in each simulated profile, only one result was obtained for each profile-by-profile comparison. In 46% of the cases a potential weak layer detected by the SSI corresponded to an observed CT failure layer.

Considering the 14 profiles where at least one CT failure layer was reported as a sudden collapse (SC) fracture, only five potential weak layers were found in the corresponding simulated profile within the given height range (POD = 0.36).

Alternatively, the simulated profiles can be considered as the reference, to check whether potential weak layers detected by the SSI in the simulated profiles have a corresponding CT failure layer in the observed profiles (c in Fig. 7). The agreement was 58% (48 cases). Again, only for 5 out of 14 cases a SC fracture was recorded.

The two methods of weak layer detection we used for the simulated profiles (d in Fig. 7) provided in 64% of the cases the same result, i.e. a layer detected by the SSI was also identified as critical by the TSA analysis. The agreement did not improve if only those potential weak layers (either detected by the SSI or the TSA) were considered that were related to SC fractures in the corresponding observed profile.

Fig. 3. Density vs. hand hardness index for faceted crystal (blue circles, N = 3471). Solid colored lines show the probability distributions for five hardness steps based on the multinomial logistic regressions.
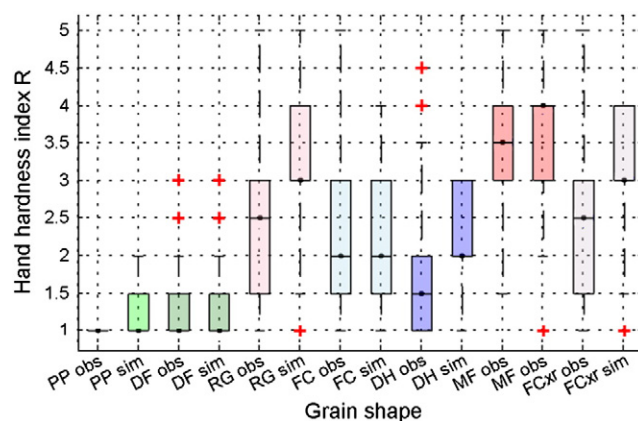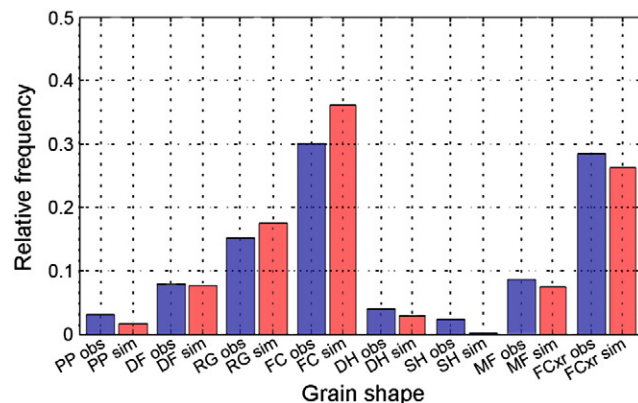

Fig. 5. Relative frequency of grain types in layers of observed (obs, blue) (N = 1790) and simulated (sim, red) profiles (N = 7926).

Table 3
Characteristics of manually observed failure layers and simulated weak layers. For each grain type the absolute number of layers (N) and their relative frequency (N_{rel}), the median grain size (E), the median hand hardness index (R) and the median number of structural variables in the respective critical range (L) are given.

Grain type	N		N_{rel}		E (mm)		R		L	
	Obs	Sim	Obs	Sim	Obs	Sim	Obs	Sim	Obs	Sim
PP	58	137	0.04	0.02	1.50	0.30	1	1	4	2
DF	142	614	0.08	0.08	0.75	0.33	1–2	1	2	2
RG	272	1388	0.15	0.18	0.38	0.51	3	3	1	1
FC	539	2861	0.30	0.36	0.88	0.87	2–3	2	2	3
DH	72	238	0.04	0.03	2.50	1.67	2	2	4	3
SH	43	17	0.02	0.00	2.00	2.00	1	1	5	5
MF	155	592	0.09	0.07	1.13	1.28	4	5	1	2
FCxr	509	2079	0.28	0.26	1.25	0.89	3	3	2	2

4.3. TSA validation

First, we evaluated whether the two methods that can be applied to detect a weak layer in the manually observed profiles, namely the TSA and the CT, provided similar results. Only 37 of 180 failure layers found with the CT were also considered critical by the TSA (a in Fig. 7, Table 4). The agreement improved from 21% to 47% (7 out 15) if only the SC fractures were considered.

The observed profiles were again considered as the reference (e in Fig. 7): for each CT failure layer we looked for a corresponding potential weak layer found by the TSA in the simulated profile. Out of the 180 observed failure layers, in 87 cases a potential weak layer detected by the TSA was found within the modeled height range (Table 4). If those CT failure layers were excluded where the fracture character was either recorded as 'break' or not recorded at all (92 observed failure layers remaining) the agreement increased from 48% to 54%. Moreover, considering only SC fractures, in 80% of the cases (12 out of 15) a simulated layer classified as potentially critical by the TSA corresponded to the observed sudden collapse failure. The three cases that had no corresponding weak layer in the simulated profile, were (1) a SC fracture located within recent snow (the primary and secondary grain type were PP and DF, respectively), (2) a collapse of the basal layers of the snowpack, and (3) a SC fracture corresponding to a persistent layer (FC).

Alternatively, the simulated profiles can be considered as the reference (f in Fig. 7), to check whether potential weak layers detected by the TSA in simulated profiles had a corresponding CT failure layer in the observed profiles. In this case, of the 233 layers identified as potential weak layers by the TSA 124 were related to CT failure layers (53% of agreement), including 12 out of the 15 cases recorded as SC fractures.
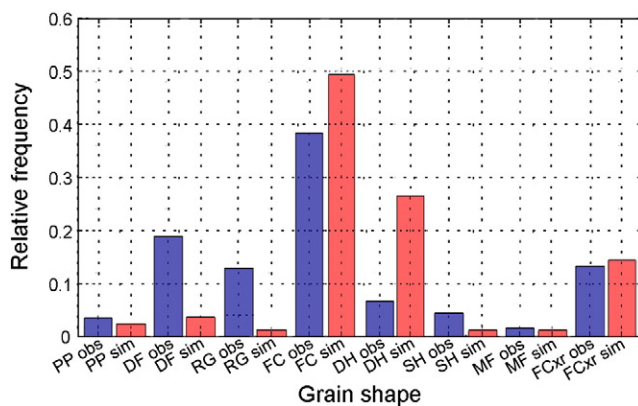

Fig. 6. Relative frequency of grain types found in CT failure layers in observed profiles (N = 180, blue bars) and potential weak layers identified by the SSI in simulated profiles (N = 83, red bars).

5. Discussion

The parameterisation of snow settlement was the most significant change in recent years made in the numerical snow cover model SNOWPACK. This change directly affects snow density and indirectly a number of other parameters such as hardness. Comparing our results to those by Schweizer et al. (2006) suggests that the recent change mainly affected the density of DH layers, producing denser DH layers than with the previous model version. Other grain types were less affected and showed a less pronounced increase in density. Simulated and observed distributions of both snow density and hardness were statistically different. However, considering the subjective manner hand hardness is estimated in the field, a difference of half a step of hand hardness index seems acceptable. We therefore did not scale the simulated densities to the observed ones in order to adjust simulated hardness. Consequently, our hand hardness index estimation is not optimized to a certain data set and will therefore not require further recalibration whenever changes to the SNOWPACK model will be made in the future.

Obviously, snow stability information derived from simulated profiles with either the TSA or the SSI can be affected by the discrepancy between simulated and observed snow density and consequently snow hardness. For assessing snow stability, hardness overestimation for layers of RG and FCxr is hardly crucial, but may be important for layers of DH. Nevertheless, a preliminary qualitative analysis indicates that the above mentioned discrepancy mainly affects layers in the lower part of the snow cover (due to the overestimation of snow density). Since most of the skier triggered avalanches are related to weak layers within the top meter of the snowpack, the hardness overestimation for the basal layers in the simulated snow profiles has no major influence on snow stability estimation.

The overall satisfactory results in weak layer detection suggest that the model in general reasonably reproduces stratigraphy characteristics. This finding is in line with previous studies. For example, Lehning et al. (2001) reported an agreement score between simulated and observed profiles of 0.65–0.75, and Monti et al. (2009) a score of 0.74. Limitations of the model in detecting potential failure layers within recent snow (i.e. PP, DF, RG) clearly exist. However, some of the discrepancy is related to the CT which is known to often indicate failure layers in recent snow that are not really critical. In addition, as the SSI is defined, it is rather unlikely that layers consisting of recent snow are identified as potential weak layers. To improve the detection of failure layers including PP and DF their grain size parameterization in SNOWPACK might need to be changed.

Within manually observed profiles, CT failure layers were used as weakness markers and compared to potential weak layers identified with the TSA (both in simulated and observed profiles) and the SSI (in simulated profiles only) or vice versa. It is well known that the compression test is not the best test and in general reveals too many weaknesses (e.g. Winkler and Schweizer, 2009). We think that detecting about half the CT failure layers by either the TSA or SSI in simulated profiles is not a poor but rather satisfactory result as the CT is not the ultimate reference for unstable weak layers — but the only one available to us. Furthermore, it cannot be expected that every CT failure layer is detectable as long as only snow stratigraphy characteristics are considered. For example, Schweizer et al. (2008) found that only weak layers where the rutschblock failed as a whole block scored as unstable in the TSA analysis. Regarding the compression test, only the micro-structural parameters of sudden collapse fractures were associated with unstable conditions (van Herwijnen et al., 2009). In particular, failure layers with no sudden fractures, near the surface are known to be less indicative. Based on the findings by van Herwijnen et al. (2009), it is suggested that primarily sudden collapse fractures can be detected with the TSA. In fact, weak layers with sudden fractures were better detected by the TSA and the SSI than other failure layers.
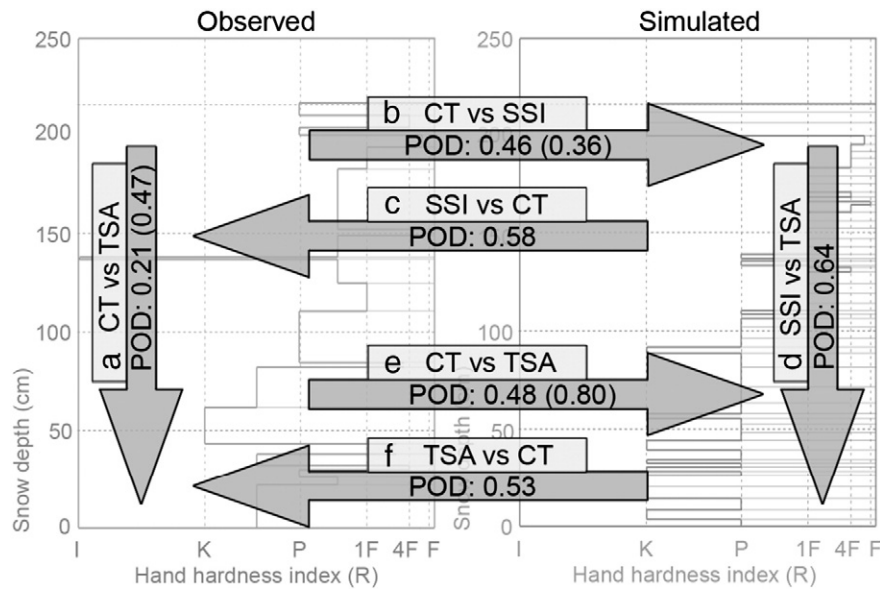
Fig. 7. Summary of the results for the comparisons described in Fig. 1. The first value in each arrow is the probability of detection (POD) considering all CT fracture types; the value in brackets represents the POD considering only the sudden collapse (SC) fracture types.

Comparing manually observed to simulated snow stratigraphy with a profile-by-profile method is in general complicated by differences in the simulated and observed parameters (e.g. different grain size definition, different layer thicknesses). Moreover, the snow pits were located a few meters from the AWS so that spatial variations in snow layering (e.g. Schweizer et al., 2008) add to the complexity of the layer matching. Though stability tests from slopes would be more indicative, matching layers observed in snow pits on slopes with layers simulated for locations of AWS would introduce an even larger uncertainty. Therefore, we only considered flat field profiles observed in close vicinity of the two AWS.

Within the manually observed profiles, CT failure layers were classified by the TSA as potentially unstable in 21% of the cases. The agreement improved to 47% if only the failure layers characterized as sudden collapse fracture (SC) were considered (a in Fig. 7) — but only few such cases existed. This confirms that the TSA is not sufficiently sensitive for this task and that snow stratigraphy alone is not suited to detect failure layers unless combined with a stability test.

Failure layers detected with the CT corresponded to the simulated potential weak layers found with the SSI in 48% of cases; the agreement did not improve if only SC fracture characters were considered (b in Fig. 7). Considering the simulated profiles as the reference, 58% of the SSI layers were related with a CT failure layer, but only 5 out of 12 SC fractures were detected (c in Fig. 7). The inability of the SSI to detect SC fractures (generally considered as the most unstable layers) might

be due to the fact that the near surface layers often have low shear strength and that the additional stress due to a skier is high (Eq. (2)). Therefore, despite the two structural parameters (Eqs. (1) and (3)) included in the SSI, the minimum value of the SSI is often found for layers close to the snow surface.

The TSA showed more promising results if combined with the simulated profiles. Using either the manual profiles as the reference, or vice versa the simulated ones, 48% of the CT failure layers were related to potential weak layers identified by the TSA within the simulated snow cover (e and f in Fig. 7). The result improved to 80% if only sudden collapse fractures were considered (e in Fig. 7). The low number of cases and the lack of objective stability information prevent a definitive assessment.

Independently of the fracture type, the TSA provided about twice as good classification results when applied to the simulated profiles compared to the observed ones (Table 4). This finding seems rather surprising and cannot easily be explained. The difference might be due to the lower resolution of the manual profiles or the fact that simulated snow stratigraphy is more consistent and less affected by observer's subjectivity. For illustration, two examples of comparison are shown in Fig. 8. In the first case (Fig. 8a), the weakness as identified by the SSI (green arrow) did not correspond to any of the two CT failure layers. However, the layer detected by the SSI was also detected by the TSA and corresponded to the lowermost potential weakness identified by the TSA in the manual profile (red arrow within the tolerance range in the observed profile). In the example shown in Fig. 8b, both CT failure layers corresponded to a simulated layer considered as critical by the TSA; on the other hand, only one of the two CT failure layers was detected as critical by the TSA applied directly to the manually observed profile itself. In both examples (Fig. 8a,b) the higher resolution of the model is obvious. As pointed out by Winkler and Schweizer (2009) different methods often do not provide the same potentially unstable weak layer. This was found for the TSA and the CT too (a and d in Fig. 7) suggesting that they are two complementary approaches at best.

Finally, the seemingly low agreement scores of about 50% have to be seen in the context of the very many layers, in fact almost 8000, to choose from, and the fact that the simulated layers are significantly thinner than the observed ones. For example, in the case of the TSA, most of the layers were properly detected as non-failure layers (i.e. 7683 out of 7926). Out of 233 identified as potential weak layers 138 were associated with a CT failure layer. However, as the simulated layers

Table 4
Detection results for CT failure layers: Agreement with potential weak layers identified by the TSA for both observed and simulated profiles per fracture type (N/A = not classified; B = break; PC = progressive compression; RP = resistant planar; SP = sudden planar; SC = sudden collapse). N is the total number of CT failure layers; TSA observed/TSA simulated is the probability of detection and the absolute number of observations (in brackets) for CT vs. TSA.

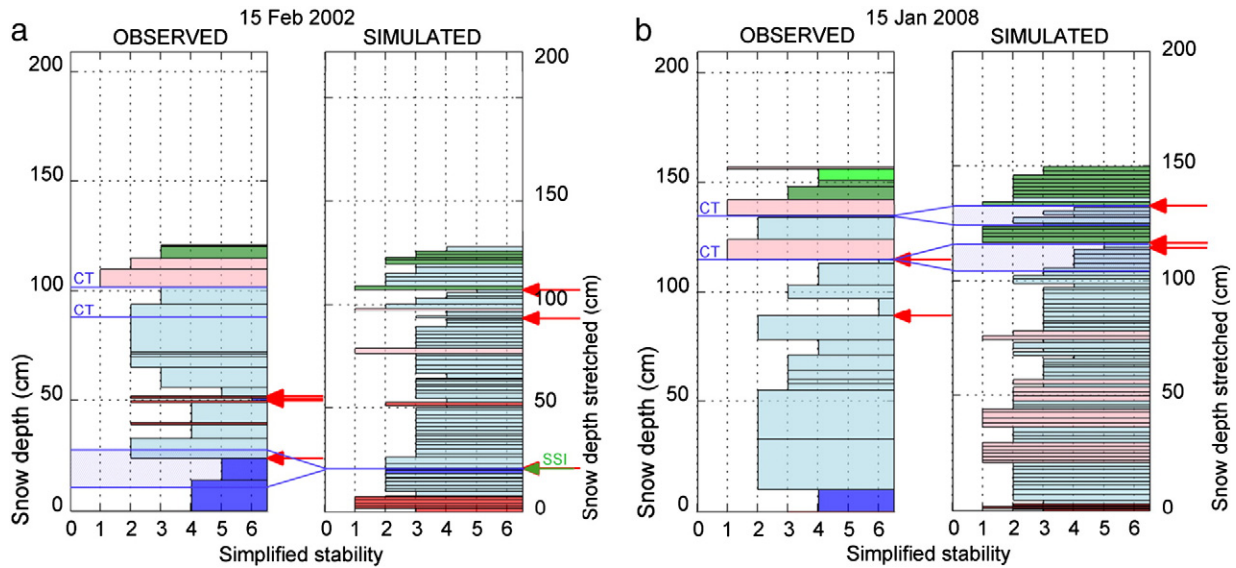|	N	TSA observed	TSA simulated
All	180	0.21 (37)	0.48 (87)
All (not incl. B, N/A)	92	0.23 (21)	0.54 (50)
PC	6	0 (0)	0.67 (4)
RP	15	0.13 (2)	0.40 (6)
SP	56	0.21 (12)	0.50 (28)
SC	15	0.47 (7)	0.80 (12)
B	41	0.10 (4)	0.29 (13)

Fig. 8. Simplified stability profiles as proposed by Monti et al. (2012). a) Comparison of the weak layer found with the SSI in the simulated profile with the corresponding manually observed profile. b) Comparison of two CT failure layers in the manually observed profile with the corresponding simulated layers.

are thin, often two potential weak layers were associated with the same CT failure layer, so that overall only 87 out of the 180 CT failures layers were detected, resulting in the abovementioned POD of 48%.

6. Conclusions

Before focusing on weak layer detection in snow profiles simulated by the numerical snow cover model SNOWPACK we validated its density parameterization and improved the hand hardness index parameterization. The newly developed hand hardness parameterization keeps hardness as categorical discrete, ordinal variable and eases the interpretation of simulated hardness profiles. As a reasonable agreement between the observed and simulated densities was obtained, we did not empirically adjust the simulated densities to improve the hardness estimation. Consequently, repeated re-calibrations due to future model improvements are not required. Simulated and observed snow characteristics were similar, suggesting that field-based methods such as the TSA can also be used for simulated snow profiles.

For validating whether SNOWPACK can indicate the existence and location of potential weak layers, we related observed snow profiles completed with at least one compression test (CT) to simulated profiles with potential weak layers that were either identified with the SSI or the TSA, two methods currently implemented in SNOWPACK. The comparisons between the methods and between observed and simulated profiles are summarized in Fig. 7. Considering the manually observed profiles as the reference profiles resulted in about 50% agreement between CT failure layers and potential weak layers detected by either the SSI or the TSA. Alternatively, if the simulated profiles were the reference profiles, the agreement with manually observed profiles was slightly higher: 58% for the SSI, and 53% for the TSA. With the TSA, the agreement clearly improved when only weak layers with sudden collapse fractures were considered.

Our findings suggest that a relation between manually observed and simulated potential weak layers exists. Therefore, deriving snow stability information from simulated snow stratigraphy for avalanche forecasting seems not only promising, but also feasible. Estimating stability for the depth of potentially weak layers detected in simulated snow stratigraphy is the next problem that needs to be tackled.

References

CAA, 2007. Observation guidelines and recording standards for weather, snowpack and avalanches. Canadian Avalanche Association (CAA), Revelstoke BC, Canada (78 pp.).

Durand, Y., Giraud, G., Brun, E., Mérindol, L., Martin, E., 1999. A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting. J. Glaciol. 45 (151), 469–484.

Fierz, C., Armstrong, R.L., Durand, Y., Etchevers, P., Greene, E., McClung, D.M., Nishimura, K., Satyawali, P.K., Sokratov, S.A., 2009. The International Classification for Seasonal Snow on the Ground. HP-VII technical documents in hydrology, 83. UNESCO-IHP, Paris, France (90 pp).

Föhn, P.M.B., 1987. The stability index and various triggering mechanisms. In: Salm, B., Gubler, H. (Eds.), Symposium at Davos 1986 — avalanche formation. Movement and effects, IAHS Publ., 162. International Association of Hydrological Sciences, Wallingford, Oxfordshire, U.K., pp. 195–214.

Geldsetzer, T., Jamieson, J.B., 2001. Estimating dry snow density from grain form and hand hardness. Proceedings International Snow Science Workshop, Big Sky, Montana, U.S.A., 1–6 October 2000. Montana State University, Bozeman MT, USA, pp. 121–127.

Jamieson, J.B., 1999. The compression test — after 25 years. Avalanche Rev. 18 (1), 10–12.

Jamieson, J.B., Johnston, C.D., 1998. Refinements to the stability index for skier-triggered dry slab avalanches. Ann. Glaciol. 26, 296–302.

Jamieson, J.B., Johnston, C.D., 2001. Evaluation of the shear frame test for weak snowpack layers. Ann. Glaciol. 32, 59–68.

Lehning, M., Fierz, C., Lundy, C., 2001. An objective snow profile comparison method and its application to SNOWPACK. Cold Reg. Sci. Technol. 33 (2–3), 253–261.

Lehning, M., Bartelt, P., Brown, R.L., Fierz, C., 2002a. A physical SNOWPACK model for the Swiss avalanche warning; part III: meteorological forcing, thin layer formation and evaluation. Cold Reg. Sci. Technol. 35 (3), 169–184.

Lehning, M., Bartelt, P., Brown, R.L., Fierz, C., Satyawali, P.K., 2002b. A physical SNOWPACK model for the Swiss avalanche warning; part II. Snow microstructure. Cold Reg. Sci. Technol. 35 (3), 147–167.

Lehning, M., Fierz, C., Brown, R.L., Jamieson, J.B., 2004. Modeling instability for the snow cover model SNOWPACK. Ann. Glaciol. 38, 331–338.

McCammon, I., Schweizer, J., 2002. A field method for identifying structural weaknesses in the snowpack. In: Stevens, J.R. (Ed.), Proceedings ISSW 2002. International Snow Science Workshop, Penticton BC, Canada, 29 September–4 October 2002. International Snow Science Workshop Canada Inc., BC Ministry of Transportation, Snow Avalanche Programs, Victoria BC, Canada, pp. 477–481.

McCullagh, P., Nelder, J.A., 1990. Generalized linear models. Chapman & Hall, New York.

Monti, F., Cagnati, A., Fierz, C., Lehning, M., Valt, M., Pozzi, A., 2009. Validation of the SNOWPACK model in the Dolomites. In: Schweizer, J., van Herwijnen, A. (Eds.), International Snow Science Workshop ISSW, Davos, Switzerland, 27 September–2 October

2009. Swiss Federal Institute for Forest, Snow and Landscape Research WSL, pp. 313–317.

Monti, F., Cagnati, A., Valt, M., Schweizer, J., 2012. A new method for visualizing snow stability profiles. Cold Reg. Sci. Technol. 78, 64–72.

Pielmeier, C., Schneebeli, M., 2003a. Developments in snow stratigraphy. Surv. Geophys. 24 (5–6), 389–416.

Pielmeier, C., Schneebeli, M., 2003b. Stratigraphy and changes in hardness of snow measured by hand, ramsonde and snow micro penetrometer: a comparison with planar sections. Cold Reg. Sci. Technol. 37 (3), 393–405.

Schweizer, J., Jamieson, J.B., 2007. A threshold sum approach to stability evaluation of manual snow profiles. Cold Reg. Sci. Technol. 47 (1–2), 50–59.

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

Schweizer, J., Wiesinger, T., 2001. Snow profile interpretation for stability evaluation. Cold Reg. Sci. Technol. 33 (2–3), 179–188.

Schweizer, J., Jamieson, J.B., Schneebeli, M., 2003. Snow avalanche formation. Rev. Geophys. 41 (4), 1016.

Schweizer, J., Bellaire, S., Fierz, C., Lehning, M., Pielmeier, C., 2006. Evaluating and improving the stability predictions of the snow cover model SNOWPACK. Cold Reg. Sci. Technol. 46 (1), 52–59.

Schweizer, J., McCammon, I., Jamieson, J.B., 2008. Snowpack observations and fracture concepts for skier-triggering of dry-snow slab avalanches. Cold Reg. Sci. Technol. 51 (2–3), 112–121.

van Herwijnen, A., Jamieson, J.B., 2007. Fracture character in compression tests. Cold Reg. Sci. Technol. 47 (1–2), 60–68.

van Herwijnen, A., Bellaire, S., Schweizer, J., 2009. Comparison of micro-structural snowpack parameters derived from penetration resistance measurements with fracture character observations from compression tests. Cold Regions Science and Technology 59 (2–3), 193–201.

Wilks, D.S., 1995. Statistical methods in the atmospheric sciences: an introduction. International Geophysics, vol. 59. Academic Press, San Diego CA, U.S.A. (467 pp.).

Winkler, K., Schweizer, J., 2009. Comparison of snow stability tests: extended column test, rutschblock test and compression test. Cold Reg. Sci. Technol. 59 (2–3), 217–226.