Geophysical Research Abstracts,
Vol. 11, EGU2009-13413, 2009
EGU General Assembly 2009

On sample size and different interpretations of snow stability datasets

M. Schirmer, C. Mitterer, and J. Schweizer
WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland (schirmer@slf.ch)

Interpretations of snow stability variations need an assessment of the stability itself, independent of the scale investigated in the study. Studies on stability variations at a regional scale have often chosen stability tests such as the Rutschblock test or combinations of various tests in order to detect differences in aspect and elevation. The question arose: 'how capable are such stability interpretations in drawing conclusions'. There are at least three possible errors sources: (i) the variance of the stability test itself; (ii) the stability variance at an underlying slope scale, and (iii) that the stability interpretation might not be directly related to the probability of skier triggering. Various stability interpretations have been proposed in the past that provide partly different results. We compared a subjective one based on expert knowledge with a more objective one based on a measure derived from comparing skier-triggered slopes vs. slopes that have been skied but not triggered. In this study, the uncertainties are discussed and their effects on regional scale stability variations will be quantified in a pragmatic way. An existing dataset with very large sample sizes was revisited. This dataset contained the variance of stability at a regional scale for several situations. The stability in this dataset was determined using the subjective interpretation scheme based on expert knowledge. The question to be answered was how many measurements were needed to obtain similar results (mainly stability differences in aspect or elevation) as with the complete dataset. The optimal sample size was obtained in several ways: (i) assuming a nominal data scale the sample size was determined with a given test, significance level and power, and by calculating the mean and standard deviation of the complete dataset. With this method it can also be determined if the complete dataset consists of an appropriate sample size. (ii) Smaller subsets were created with similar aspect distributions to the large dataset. We used 100 different subsets for each sample size. Statistical variations obtained in the complete dataset were also tested on the smaller subsets using the Mann-Whitney or the Kruskal-Wallis test. For each subset size, the number of subsets were counted in which the significance level was reached. For these tests no nominal data scale was assumed. (iii) For the same subsets described above, the distribution of the aspect median was determined. A count of how often this distribution was substantially different from the distribution obtained with the complete dataset was made. Since two valid stability interpretations were available (an objective and a subjective interpretation as described above), the effect of the arbitrary choice of the interpretation on spatial variability results was tested. In over one third of the cases the two interpretations came to different results. The effect of these differences were studied in a similar method as described in (iii): the distribution of the aspect median was determined for subsets of the complete dataset using both interpretations, compared against each other as well as to the results of the complete dataset. For the complete dataset the two interpretations showed mainly identical results. Therefore the subset size was determined from the point at which the results of the two interpretations converged. A universal result for the optimal subset size cannot be presented since results differed between different situations contained in the dataset. The optimal subset size is thus dependent on stability variation in a given situation, which is unknown initially. There are indications that for some situations even the complete dataset might be not large enough. At a subset size of approximately 25, the significant differences between aspect groups (as determined using the whole dataset) were only obtained in one out of five situations. In some situations, up to 20% of the subsets showed a substantially different distribution of the aspect median. Thus, in most cases, 25 measurements (which can be achieved by six two-person teams in one day) did not allow to draw reliable conclusions.