

VARIATIONS IN INDIVIDUAL DANGER LEVEL ESTIMATES WITHIN THE SAME FORECAST REGION

Frank Techel*^{1,2}, Lukas Dürr¹ and Jürg Schweizer¹

¹WSL Institute for Snow and Avalanche Research SLF, Davos

²University of Zurich, Department of Geography, Zurich

ABSTRACT: Forecasters in many public avalanche forecast operations rely on field observations and danger (or stability) assessments by experienced and well-trained observers. However, available field observations are often sparse, represent local conditions and may even be contradictory. We addressed these issues by comparing the local danger level estimates between different observers within the same forecast region (approx. 225 km²). 1120 local danger assessment pairs within the same region were compared. The agreement between the danger estimates of different observers in the same forecast region was 79%. Differences were likely due to local variations in avalanche conditions, but also observer-specific biases existed. Despite challenges in data interpretation due to locally variable avalanche conditions, poor spatial coverage and observer-specific characteristics, danger assessments based on field observations are key information for avalanche forecasting. However, such danger level estimates should always be considered in combination with other information related to the avalanche situation.

KEYWORDS: avalanche forecasting, avalanche danger level, observer-specific bias

1. INTRODUCTION

The day-to-day review of the avalanche forecast of the previous day is the starting point of the forecasting process on the current day – at least this is the approach followed by the Swiss avalanche forecasting service at SLF. However, as avalanche danger cannot be measured, a true verification of the forecasted avalanche danger level is not possible (Föhn and Schweizer, 1995). Information related to snow instability – such as recent avalanches, signs of instability or stability tests – provide the most direct information on the avalanche conditions (e.g. McClung and Schaerer, 2006). However, in an operational setting this kind of information is often not readily available, particularly at lower danger levels (no or minor avalanche activity and no direct signs of instability) or in poor weather conditions (no or only limited observations). Thus, the day-to-day operational review of the regional avalanche forecast has to make use of numerous other information sources, such as weather and snowpack observations less directly related to the danger level. In addition, as is the case in Switzerland, specifically trained observers provide local danger level estimates for the current day.

Although local avalanche danger assessments are subjective interpretations based on field observations, they are considered fairly accurate estimates of the avalanche danger (Schweizer, 2010). While SLF forecasters consider the quality of local danger level estimates from experienced observers to be high, uncertainties in data interpretation may be caused by spatially variable avalanche conditions (e.g. Schweizer et al., 2003), but there is also the chance that observational biases exist.

As is the case with any relevant information source used in the process of avalanche forecasting, it is also important to know the limitations of a parameter such as the estimated local avalanche danger level. This estimated parameter probably incorporates even more previous local knowledge and experience than most other routinely reported snow and avalanche observations.

Therefore, our main aim was to explore the accuracy and limitations of local avalanche danger estimates. We compared individual estimates by different observers in the same small forecast region to assess if significant observer-specific biases exist.

2. DATA

Observers of the Swiss avalanche forecasting service with sufficient experience and view into avalanche terrain provide an estimate of the avalanche danger level together with their observations (SLF, 2013). The danger level estimate is based on the 5-level European avalanche danger

* *Corresponding author address:*

Frank Techel, WSL Institute for Snow and Avalanche Research SLF, Flüelastrasse 11, 7260 Davos Dorf, Switzerland;
email: techel@slf.ch

scale (EAWS, 2006). In addition, at danger level 3 (Considerable), observers may indicate whether or not they expect natural avalanches. The local estimate of the danger level (D_{LN}) should integrate all available information, including not just the observations from the day of observation, but also prior knowledge concerning the development of the snowpack during the winter or information from third parties. To ensure consistent and high quality feedback, all observers are regularly trained.

The observers locally assess the avalanche danger; the area considered is the area of observation during the day in the field or in the ski area, or the area that can be seen from the observation point in the valley floor. For a typical ski touring day (in Canada), Jamieson et al. (2008) estimated this area to be approximately 10 km². While estimating the danger level, the type of avalanche (dry- or wet-snow) is a main criterion to be considered. The estimated danger level for dry-snow avalanches should reflect the current situation and is therefore a local nowcast, while for wet-snow avalanches the highest expected danger level during the day is reported. Furthermore, the observers also indicate the most critical slope aspects and elevation bands (danger rose).

Measurements and field observations are transmitted by observers predominantly using two channels: the IFKIS web-platform (established in 2001, Bründl et al., 2004), which might be compared to the InfoEX in Canada (CAA, 2016; Haegeli et al., 2014), and the mAvalanche mobile App, introduced in 2008 (Suter et al., 2010).

We analyzed avalanche danger assessments reported between 11 am and 10 pm through either the IFKIS or the mAvalanche network in the winters between 2008-2009 and 2015-2016. We considered all local danger assessments related to dry-snow avalanches in the Swiss Alps for days when an avalanche forecast was issued in the morning. We were primarily interested in danger level estimates related to a day of backcountry touring or off-piste skiing. Therefore, we excluded danger assessments based on observations made in ski areas or from the valley floor. Furthermore, at least two estimates for a given forecast region had to be reported. In total, we explored danger assessments on 456 different days from 62 different forecast regions resulting in 1120 pairs of estimates.

The mAvalanche network was the primary data source used (76% of danger assessments vs.

24% from IFKIS). Many of the observers considered in this analysis were professional mountain guides, but also include SLF avalanche forecasters and other SLF staff members.

3. METHODS

We were interested mainly in the variations of the local danger level estimate D_{LN} between observers within the *same* region. The *same* region we defined by using the smallest spatial unit used in the Swiss avalanche bulletin. These spatial units, the forecast regions, divide the Alpine forecast area (26'400 km²) into 117 regions with a mean size of 225 km². While a danger level is given for the whole forecast area, i.e. each of the 117 forecast regions, the forecast regions are not explicitly used in the avalanche bulletin, since they are aggregated to larger regions with similar avalanche conditions (Pielmeier and Winkler, 2012; Ruesch et al., 2013).

We calculated the difference between D_{LN} between observers using the integer values assigned to the five danger levels. If the difference was equal to zero, we called it an agreement, else it was a disagreement. The rate of disagreement (R_d) was therefore the ratio of disagreements to all comparisons. The mean D_{LN} variation (Var_{mean}) was calculated for each observer as follows:

$$Var_{mean} = \frac{N_{high} - N_{low}}{N_{high} + N_{equal} + N_{low}}$$

where N_{high} , N_{equal} and N_{low} are the number of higher, equal or lower D_{LN} , respectively, reported by others in the same region.

Furthermore, we explored whether the disagreements were randomly distributed or a significant bias towards lower or higher disagreements existed. To this end, we calculated the proportion of equally distributed disagreements ($N_{high} = N_{low}$), and the unbalanced disagreements (equivalent to Var_{mean}). While the first, the equally distributed disagreements may be interpreted as random, a significant proportion of unbalanced disagreements Var_{mean} may indicate an observer-specific bias. As an example, observers A to D in Table 1 have unbalanced disagreements to others, while observer E's disagreements were equally often higher and lower. We tested whether N_{high} and N_{low} were significantly different than an equal distribution of N_{high} and N_{low} using the chi-square based non-parametric proportion test (R Core Team, 2016). The calculation of the p -value is sensitive to both the absolute number of disagreements as well as the absolute number of higher or lower

values. Again, some examples, shown in Table 1, may highlight this: While A and C have the same mean variation Var_{mean} (+0.2), this is only significant for C, where the sample is based on a greater number of disagreements (40 for C vs. 20 for A).

In contrast, A and B have the same absolute disagreement rate R_d , but differ in Var_{mean} (+0.2 for A vs. +0.32 for B). The bias would be considered significant for B only.

Table 1: Examples (A to E) highlighting the role of the absolute number of higher and lower disagreements (N_{high} and N_{low}) when testing whether a bias was significant or not. N is the total number of comparisons, N_{equal} the number of same danger level estimates, R_d the proportion of disagreements, Var_{mean} the mean difference in danger level estimate, and p denotes the level of significance.

Observer	N	N_{high}	N_{equal}	N_{low}	R_d	Var_{mean}	p
A	50	15	30	5	0.4	+0.2	0.19
B	50	18	35	2	0.4	+0.32	0.02
C	100	30	60	10	0.4	+0.2	0.04
D	100	20	70	10	0.3	+0.1	0.29
E	100	15	70	15	0.3	0	1

4. RESULTS

1120 local danger rating pairs were analyzed. In 90% of the cases, where the exact location was known, the location of the observers was 12 km or less from each other (mean distance 5.5 km). In 79% the local danger level estimate D_{LN} agreed, 10% were balanced disagreements ($N_{high} = N_{low}$) and 11% were unbalanced disagreements.

The D_{LN} agreement rate was 77% if the assessment was reported via different networks (mAvalanche and IFKIS) or via IFKIS. In these cases, observers were unable to see the D_{LN} others had reported previously. This compares to a slightly higher agreement rate in D_{LN} in the mAvalanche network (81%, $p > 0.05$), a network which was established with the intention to also allow a convenient information exchange between users.

Exploring observer-specific biases, considerable variations became obvious (Fig. 1). Some observ-

ers estimated D_{LN} more often lower or higher compared to others in the same region. Of the 32 observers with more than 20 D_{LN} comparisons to others, this observer-specific bias was significant for 6 of 32.

Two examples are highlighted in Figure 1:

- Observer A: 39 out of 149 (25%) estimates disagreed (x-axis). However, the difference was almost as frequently higher ($N_{high} = 21$) or lower ($N_{low} = 16$) resulting in an almost balanced value of $Var_{mean} = -3\%$ (y-axis).
- Observer B: 34 out of 132 (26%) estimates disagreed. Out of these 34 cases, 25 were lower and 9 were higher resulting in Var_{mean} of -12% . This variation was significantly different ($p < 0.01$) compared to an equal distribution of N_{high} and N_{low} .

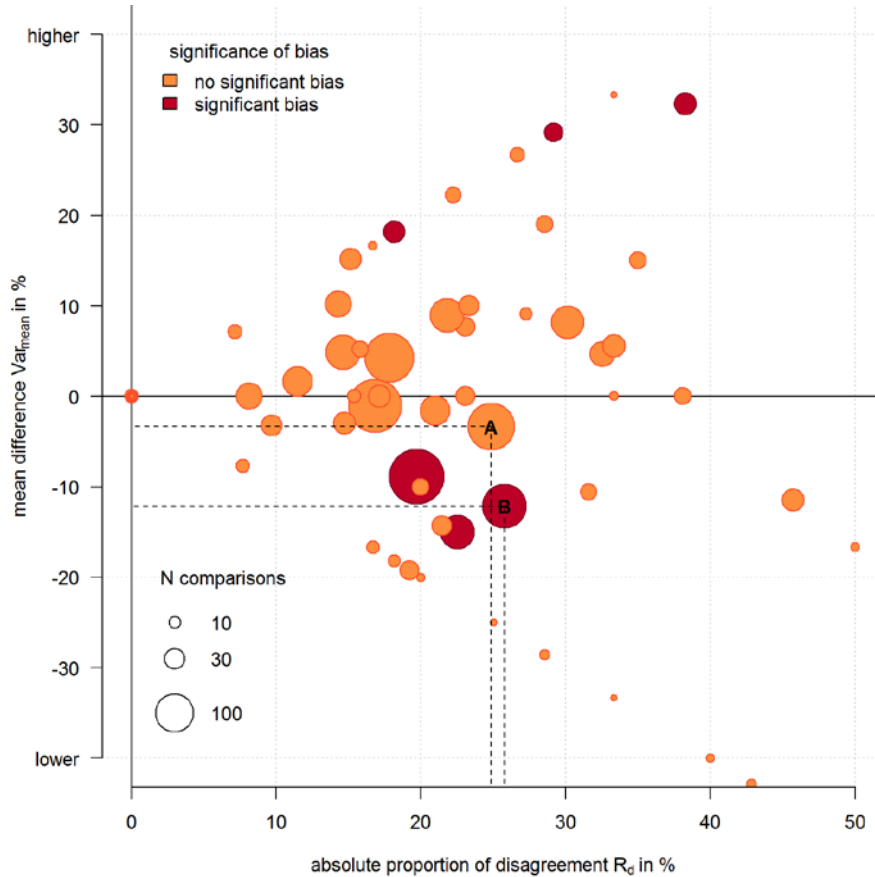


Figure 1: Comparison between local danger estimates by different observers within the same small forecast region. The size of the circles corresponds to the number of comparisons, the color to whether the bias was significant or not. The x-axis shows the absolute proportion of disagreements R_d , and the y-axis the mean difference in danger level estimates, or in other words, the unequally balanced proportion of higher and lower D_{LN} disagreements. Two examples A and B are explained in the text.

5. DISCUSSION

5.1 *Variations in local danger level estimates*

21% of the local danger level estimates disagreed within the same region, and about half of these disagreements were, considering observers individually, unbalanced towards either higher or lower danger level estimates. This highlights that at least some observers had a significant tendency towards consistently lower or higher D_{LN} than others in the same area. However, only some of these variations can be explained by such observer-specific characteristics. Further explanations may be related to spatially differing avalanche conditions. It is possible that some observations were made in relatively frequently tracked terrain (for instance close to ski areas) and others from backcountry touring regions, but also that snow-

pack characteristics varied within a relatively small region (e.g., Schweizer et al., 2003). And finally, there may be situations when the avalanche danger is somewhere between two danger levels, but observers have to decide on one level in their reporting form.

5.2 *Interpretation of several corresponding local danger estimates*

Local danger level estimates by experienced professionals are considered fairly accurate and an important source of information for forecasters at SLF. Two, or several, corresponding danger level estimates by experienced professionals from the same forecast region may thus be seen as a relatively reliable estimate of the regional danger level. However, a detailed analysis in hindsight may sometimes show that even on days when several

observers agree the avalanche danger level was in fact different. We illustrate this observation with one example (31 December 2014), when the avalanche situation was likely more critical than anticipated by the observers in the field, but as well as by the avalanche warning service. On that day the visibility was poor until late in the afternoon in the region of Davos. Field observations and danger level estimates were reported by two mountain guides and two SLF forecasters. All four described frequent danger signs such as whumpfs or shooting cracks at elevations above tree line, and three of them remotely triggered small slab avalanches. Based on their observations, all four estimated the danger level as 3 – in agreement with the forecast. Two of the observers expected natural avalanches, the other two did not. These observations are typically associated with a danger level 3, but may also correspond to a danger level 4 (Schweizer, 2010). The next day, when visibility was good, numerous small, medium and even some large natural avalanches, but also human-triggered and explosive controlled avalanches were observed. The re-analysis of the situation showed that the last days before New Year had been one of the periods with the highest avalanche activity during the last years. Considering all information, in hindsight the avalanche warning service verified the danger level as 4, and thus higher than all four field assessments.

5.3 Practical implications for avalanche forecasting services

Timely and accurate field observations and danger level estimates by trained professionals are an important source to check the accuracy of the forecasted avalanche danger level. However, these local, and to a certain extent subjective, estimates must be interpreted in the context of an often much larger regional scale. This can be a challenging task, particularly if field observations are spatially sparse, and sometimes even contradictory.

The observed differences in local danger estimates highlight some important points.

There is a need to reduce observer-specific bias'. This can only be achieved by regularly training (even experienced) observers, not just in how to carry out observations and estimates in the field, but also in correctly using the standardized snow and avalanche terminology.

Observation networks and information exchange platforms provide the means to report standardized snow, weather and avalanche information. However, they also serve as a valuable information source providing current and relevant information on snow and avalanche conditions for observers and avalanche professionals. Yet, particularly in a setting like in Switzerland, where observation density may be higher than in many other regions, it is debatable whether individual stability assessments or danger level estimates, should be made accessible to others, or if access should be limited to more general weather, snow and avalanche observations to avoid mutual influence.

In our study, in one out of five cases, there was a disagreement between danger level estimates. This also means that there is a limit on how reliably forecasters can verify the regionally forecasted danger level using single local danger level estimates, even at relatively small scales of 200 km². In particular for large and data-sparse regions, an operational verification using field observations therefore seems almost impossible.

CONCLUSIONS AND OUTLOOK

We have analyzed estimates of local avalanche danger reported by trained observers for the same forecast region. While the agreement between individual estimates was relatively high (79%), we sometimes noted an observer-specific reporting bias highlighting the importance that even experienced observers and professionals should regularly attend observer training courses and be trained in the use of common standards. The disagreement rate of 21% also shows the difficulty of assessing the avalanche situation, even in the field, and describing it with a single danger level.

Local danger level estimates by experienced professionals and observers are a very valuable source for the operational verification of the avalanche forecast. However, they should always be complemented with other field observations allowing the forecaster to obtain a more comprehensive picture on the local avalanche situation. Also, given the agreement rate between individual observers, it seems questionable whether the accuracy of the avalanche forecast can be higher than the agreement rate.

Future work will explore whether different field observations may provide an explanation why danger level estimates differed, and whether local

danger level estimates correspond with the forecasted danger level.

REFERENCES

- Bründl, M., Etter, H.J., Steiniger, M., Klingler, C., Rhyner, J. and Ammann, W., 2004. IFKIS - a basis for managing avalanche risk in settlements and on roads in Switzerland. *Nat. Hazards Earth Syst. Sci.*, 4(2): 257-262.
- CAA, 2016. InfoEX - Industry standard for an extraordinary industry. <https://www.avalancheassociation.ca/?page=InfoEx> (accessed 12 Aug 2016).
- EAWS, 2006. European Avalanche Danger Scale. http://www.avalanches.org/eaws/en/main_layer.php?layer=basics&id=2 (accessed 12 Aug 2016).
- Föhn, P.M.B. and Schweizer, J., 1995. Verification of avalanche danger with respect to avalanche forecasting. In: F. Sivardière (Editor), *Les apports de la recherche scientifique à la sécurité neige, glace et avalanche. Actes de Colloque, Chamonix, 30 mai-3 juin 1995*. ANENA, Grenoble, France, pp. 151-156.
- Haegeli, P., Obad, J., Harrision, B., Murray, B., Engblom, J. and Neufeld, J., 2014. InfoEX™ 3.0 - advancing the data analysis capabilities of Canada's diverse avalanche community. In: P. Haegeli (Editor), *Proceedings ISSW 2014. International Snow Science Workshop, Banff, Alberta, Canada, 29 September - 3 October 2014*, pp. 910-917.
- Jamieson, J.B., Campbell, C. and Jones, A., 2008. Verification of Canadian avalanche bulletins including spatial and temporal scale effects. *Cold Reg. Sci. Technol.*, 51(2-3): 204-213.
- McClung, D.M. and Schaerer, P., 2006. *The Avalanche Handbook*. The Mountaineers Books, Seattle WA, U.S.A., 342 pp.
- Pielmeier, C. and Winkler, K., 2012. The new and improved Swiss avalanche bulletin. *The Avalanche Review*, 31(2): 15.
- R Core Team, 2016. *A Language and Environment for Statistical Computing*. Version 3.3.1 (2016-06-21). R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/> (accessed 12 Aug 2016).
- Ruesch, M., Egloff, A., Gerber, M., Weiss, G. and Winkler, K., 2013. The software behind the interactive display of the Swiss avalanche bulletin. In: F. Naaim-Bouvet, Y. Durand and R. Lambert (Editors), *Proceedings ISSW 2013. International Snow Science Workshop, Grenoble, France, 7-11 October 2013*. ANENA, IRSTEA, Météo-France, Grenoble, France, pp. 406-412.
- Schweizer, J., 2010. Predicting the avalanche danger level from field observations, *Proceedings ISSW 2010. International Snow Science Workshop, Lake Tahoe CA, U.S.A., 17-22 October 2010*, pp. 162-165.
- Schweizer, J., Kronholm, K. and Wiesinger, T., 2003. Verification of regional snowpack stability and avalanche danger. *Cold Reg. Sci. Technol.*, 37(3): 277-288.
- SLF, 2013. *SLF-Beobachterhandbuch: Regionale Beobachter, Hangprofiler, Geländebeobachter*. WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland.
- Suter, C., Harvey, S. and Dürr, L., 2010. mAvalanche - Smart avalanche forecasting with smartphones, *Proceedings ISSW 2010. International Snow Science Workshop ISSW, Lake Tahoe CA, U.S.A., 17-22 October 2010*, pp. 630-635.