ASSESSING THE COLORADO AVALANCHE INFORMATION CENTER'S BACKCOUNTRY AVALANCHE FORECASTS

Spencer Logan[1], Ethan Greene[1]
[1] *Colorado Avalanche Information Center, Denver, CO, USA*

ABSTRACT: Understanding the quality of a prediction is an integral part of improving any forecasting process. The Colorado Avalanche Information Center (CAIC) selected a set of public backcountry avalanche forecasts issued between November 2017 and April 2022 for verification. The forecasts came from continuous periods of one or more months and were from locations with a significant volume of observations. An individual assessor determined the avalanche danger and avalanche problems for a single day. Each assessor used the avalanche danger and problems from the day before their assigned period and a dataset of observations that included records from before and after each assessed day. Each assessment consisted of an avalanche danger rating for up to three elevation bands, up to three avalanche problems, a critical factor, and a confidence level. We calculated the avalanche activity index (AAI) for each day and location. Here we present the results of assessing three seasons and almost 1000 forecast days. The forecast and assessed avalanche danger rating matched 84% of the days, with a similar proportion of days where the assessment was higher or lower than the forecast. The difference between forecast and observed weather was the critical factor for many days when the assessment was lower than the forecast danger. The AAI was consistently and significantly higher on days when the assessed danger was higher than the forecast danger compared to days when the assessed danger matched or was lower than the forecast danger.

KEYWORDS: forecasting, verification, avalanche danger

1. INTRODUCTION

Verifying a forecast is an important part of any predictive process. It allows us to understand how well forecasts represent reality, examine our forecast process, and identify biases that affect them. A skilled prediction must be better than a good guess, and without a systematic way of evaluating forecasts, forecasters cannot understand or improve their performance.

Evaluating avalanche forecasts is a challenging task. Public safety groups in North America and Europe issue avalanche forecasts for large areas and rate the danger of avalanches on a 5-level scale that combines the distribution of events, the likelihood of an event, and the size of the expected event (Statham et al., 2010; EAWS, 2023). Avalanche safety groups have worked to verify regional, categorical forecasts (Elder and Armstrong, 1987; Föehn and Scweizer, 1995; McClung, 2000; Schweizer et al., 2003; Jamieson et al., 2007; Sharp, 2014; Techel, 2017; Statham et al., 2018b; Schweizer et al., 2020; Techel, 2020). The number of factors combined in each level of these scales and the rarity of avalanche events makes for a difficult verification problem but does not decrease the importance of understanding the performance of these predictions (Ebert and Milne, 2022).

This paper focuses on our recent efforts to examine the forecast *quality*, the degree to which a forecast corresponds to the actual occurrence (Murphy, 1993), of the CAIC's public backcountry avalanche forecasts. We evaluated the forecasts by comparing them to hindsight assessments. We created the hindsight assessments using observed avalanche occurrences, snowpack properties, and weather

* *Corresponding author address:*
Spencer Logan, Colorado Avalanche Information Center, 1313 Sherman Street Denver, CO 80203 USA; tel: +1 303-499-9650
email: spencer.logan@state.co.us

conditions. We relied on the adage that "the benefit of hindsight" provides us with a perspective to evaluate, understand, and improve our predictions of regional backcountry avalanche conditions.

2. METHODS

2.1 Study Site

The state of Colorado is located in the central Rocky Mountains of North America and is characterized by a continental snow climate (Mock and Birkeland, 2000). The CAIC issued public backcountry avalanche forecasts daily from November 1 until May 30, for ten regional forecast zones until the winter of 2022-23 (Figure 1). These ten zones include a total area of approximately 65,000 km^2, with the area of each forecast zone varying from about 3900 km^2 to about 11,700 km^2.

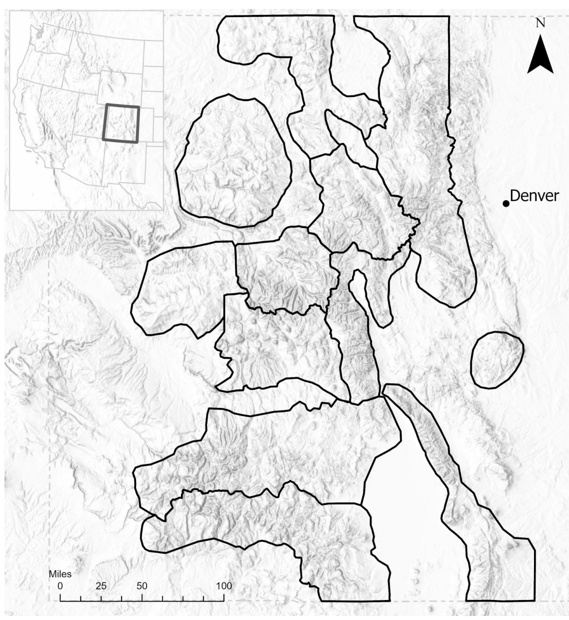

Figure 1. Map of the study areas in western Colorado, USA. The CAIC's ten forecast zones used for the assessments are outlined.

2.2. Backcountry Forecast Process

The CAIC's public backcountry avalanche forecasts include an avalanche danger rating for up to three elevation bands (above, near, and below treeline). The greatest danger of the elevation dangers is the overall, or Tier 1 (T1), danger. The forecasts include up to three avalanche problems (avalanche problem type, location, likelihood of trigger, avalanche size). Danger ratings and avalanche problems are based on the Conceptual Model of Avalanche Hazard (Statham et al., 2018a). A team of three or more staff members reviews current avalanche and weather forecasts and recent observations. The team comes to a consensus for each element of the backcountry forecast and then creates a suite of forecast products for the public.

2.3 Assessment Process

This paper focuses on assessments we conducted for the winters of 2017-18, 2020-21, and 2021-22. The assessment process was the same in all winters, but the selection of forecast sets to assess differed. In 2017-18, we created three sets of assessments. Each set covered two consecutive months. Each day of the period, one of the ten forecast zones was randomly selected. We provided assessors with the previous day's forecast for the relevant zone. Two different assessors examined each set.

For the winter of 2020-21 and 2021-22, each assessment set was two consecutive months for the same forecast zone. The zones selected were data-dense, with large numbers of avalanche observations, field reports, and weather observations. We provided assessors with an initial danger rating and avalanche problem for the day prior to the start of the assessment period. At least two assessors examined each set.

We changed zone selection criteria based on forecaster input after the 2017-18 assessments. Forecasters felt that the randomly selected zones required too much cognitive switching, particularly in the CAIC's data-sparse forecast zones. Assessing a single day's forecast in a data-sparse zone often required researching the entirety of the season's weather and snowpack development, a much more complicated process than our forecaster's typical process of tracking the daily development of the snowpack. The 2020-21 and 2021-22 sets required much less effort from the assessors because they did not require switching between locations.

Our dataset was composed of 1170 assessments. Each assessment included day-one danger ratings at three elevation bands and up to three day-one avalanche problems (avalanche problem type, likelihood, and size), similar to the CAIC's public backcountry avalanche forecasts. The assessments also include a rating of agreement and evidence to determine forecaster confidence (Mastrandrea et al., 2010). The assessors chose a critical factor adapted from Lazar et al. (2016). Our analysis here focuses on danger ratings, avalanche occurrence, and critical factors.

2.4 Avalanche Activity Index

We calculated the Avalanche Activity Index (AAI, Schweitzer et al., 2003) for the zone on each assessment day. The AAI is the sum of all recorded avalanches scaled by Destructive Size, with D1 through D4 given a value of 0.01, 0.1, 1, and 10, respectively. We calculated an overall AAI from all avalanches in the zone for a day and AAI for each elevation band. Our experience at the CAIC is similar to that reported by Schweiter (2003), with AAI providing a valuable indication of avalanche danger, but not sufficient by itself to distinguish between danger rating levels.

3. RESULTS AND DISCUSSION

3.1 Agreement and Root Mean Square Error

The root mean square error (RMSE) compares predicted and observed values. The smaller the RMSE, the closer the two values are. Table 1 shows the agreement between the forecast and assessed dangers, and the RMSE. The T1 dangers show good agreement. The RMSE suggests a discrepancy of less than one danger rating, and in all but one instance, the assessed dangers were within one category above or below the issued danger.

Observational bias may contribute to the larger RMSE below treeline. Observing avalanches below treeline from a distance is more challenging, so avalanche activity may be underreported despite CAIC forecasters collectively spending more time traveling below treeline. Ten percent of the assessments indicated that the forecast danger was one rating too high below treeline, which suggests forecasters may be reluctant to have a large difference in danger ratings between the higher elevations and below treeline.

Table 1. The percent of agreement between assessed and issued avalanche danger and the RMSE for the T1 and elevation bands.

Category	*Agreement*	*RMSE*
T1	84%	0.403
Above treeline	81%	0.378
Near treeline	81%	0.383
Below treeline	79%	0.416

3.2 Tier 1 Contingency Tables and AAI

Table 2 is the contingency table for T1 danger ratings. The Level 5, Extreme danger rating is omitted from all contingency tables shown because that rating was never used during the forecast period or in assessments. The assessed and forecast T1 dangers matched 84% of the time, with nearly equal distributions of over and under forecasts. Four percent of the assessments indicate over-forecasting when a Considerable danger was issued, and four percent when a Moderate danger was issued. About ten percent of the assessments indicated under-forecasting when Moderate or Considerable danger was issued.

Days where the assessments indicated under-forecasting had much higher AAI than days where the forecast and assessments matched or indicated over-forecasting (Figure 2). Days where the T1 danger was over-forecast had similar median AAI but skewed to lower values, indicating fewer avalanches overall in this pool. There was a single day when the issued T1 danger was two ratings higher than the assessed danger. On that day, there was little avalanche activity reported, giving a low AAI.

Table 2. A contingency table of assessed (rows) and issued (columns) T1 danger ratings. The central diagonal shows forecasts where dangers matched, in 84% of the assessments.

Assessed T1 Danger				
High	0	0	5	13
Cons.	0	54	192	7
Moderate	31	550	48	0
Low	228	41	1	0
Issued T1 Danger	*Low*	*Moderate*	*Cons.*	*High*

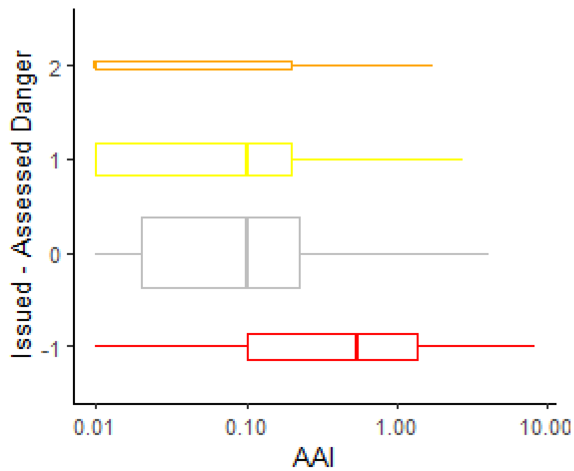

Figure 2. Boxplots of AAI of issued compared to assessed T1 dangers. Box widths are scaled by the data. Under-forecast (assessed danger greater than issued) days had a much greater AAI, while over-forecast days had a much lower AAI.

The most frequent source of discrepancy between issued and assessed danger tends to come at the end of avalanche cycles, with the assessments indicating the avalanche danger could have decreased by a rating sooner than the forecasts. Another common source of discrepancy was a single-day increase in avalanche activity that assessments indicated tended to be under forecast. Much of that avalanche activity was related to new snow issues (e.g., Storm Slab, Wind Slab avalanche problem types), while the primary avalanche problem was persistent slab avalanches that did not show a commensurate increase on those days. These discrepancies highlight the challenge inherent in a forecast: determining the actual regional avalanche activity in advance in broad, probabilistic categories. The CAIC forecasts also focus on the avalanches most dangerous to backcountry travelers. In Colorado's continental snowpack climate, those avalanches break on persistent weak layers. The issued danger may be weighted heavily by the persistent slab avalanche problem and reflect less the smaller avalanches in the upper snowpack that can lead to spikes in AAI.

3.3 T1 Danger and Critical Factors

Assessors could choose one critical factor from a list we developed based on Lazar et al.'s 2016 study of forecast consistency (Table 3). The number of avalanches was the critical factor in a quarter of assessments, indicating the importance that assessors placed on the avalanche activity. Precipitation was the critical factor selected in almost a quarter of assessments. It was the factor in the single assessment where the danger was assessed two ratings lower than issued. In that instance, the forecast snowfall amount was much greater than the amount of snow that actually fell, and the issued danger rating was tied to the worst case, most dangerous precipitation forecast. Snowpack structure was selected in almost a quarter of the assessments, indicating the importance of persistent weak layers in the assessor's evaluation of avalanche danger.

The selection of critical factors by assessed T1 danger showed several intriguing patterns (Table 4). The number of avalanches and precipitation were both key factors and did not vary much by assessed T1 danger rating. Snowpack structure was selected less at greater assessed danger ratings. Avalanche size became increasingly important as the danger increased. Size is one factor determining the difference between Considerable and High danger, and is weighted heavily in the CAIC's forecast process. Avalanche Problem Type (Statham et al., 2018a) captures the hazard avalanches pose to travelers and was selected more as the assessed dangers increased.

Table 3. Critical factors and the number of times assessors selected them, based on agreement with issued and assessed dangers (n=1170).

Critical Factor	*Under forecast*	*No difference*	*Over forecast*	*Over forecast by two danger ratings*	*Total times Critical Factor selected*	*Percent of times Critical Factor Selected*
Number of avalanches	30	229	30	0	289	25%
Precipitation	20	232	17	1	270	23%
Snowpack structure	14	234	21	0	269	23%
Avalanche Problem Type	5	57	0	0	62	5%
Avalanche size	4	43	5	0	52	4%
Old snow surface	1	39	8	0	48	4%
Temperatures	7	31	7	0	45	4%
Winds	9	23	2	0	34	3%
Test results	0	16	1	0	17	1%
(none selected)	0	79	5	0	84	7%

Table 4. The percentage of times each critical factor was selected, by assessed T1 danger rating. Each column of danger totals 100%.

	Assessed T1 Danger			
Critical Factor	*Low*	*Moderate*	*Cons*	*High*
Number of avalanches	21%	25%	28%	22%
Precipitation	27%	19%	28%	22%
Snowpack structure	30%	24%	12%	11%
Avalanche Problem Type	0%	7%	8%	11%
Avalanche size	3%	4%	6%	17%
Old snow surface	5%	5%	2%	0%
Temperatures	1%	5%	5%	0%
Winds	0%	3%	4%	6%
Test results	1%	2%	0%	0%

3.3 Elevational Avalanche Danger Contingency Tables and AAI

The elevational avalanche dangers showed a slightly poorer correspondence with the assessed dangers than the T1 dangers did (Table 5). Above and Near treeline dangers matched 81% of the time, while Below treeline dangers matched 79% of the time. The assessments indicated that Below treeline dangers were over-forecast 14% of the time–about twice the rate of under-forecasting. Like the RMSE, observational bias and forecaster decisions may contribute to the larger differences for Below treeline dangers, particularly in cases of over-forecasting.

Reported AAI by elevation band (Figure 3) shows the importance of avalanche activity in the assessments, similar to the selection of critical factors. For most days when forecasts

TABLE 5. Contingency tables of assessed (rows) and issued (columns) for avalanche dangers at each elevation band.

Above Treeline

Assessed T1 Danger				
High	0	2	10	12
Cons.	1	72	224	19
Moderate	38	416	44	0
Low	195	26	0	0
Issued T1 Danger	*Low*	*Moderate*	*Cons.*	*High*

Near Treeline

Assessed T1 Danger				
High	0	1	10	12
Cons.	1	80	203	13
Moderate	52	493	42	0
Low	241	22	0	0
Issued T1 Danger	*Low*	*Moderate*	*Cons.*	*High*

Below Treeline

Assessed T1 Danger				
High	0	2	9	7
Cons.	1	76	111	5
Moderate	80	418	36	0
Low	392	33	0	0
Issued T1 Danger	*Low*	*Moderate*	*Cons.*	*High*

and assessments matched, the AAI was similar at all three elevation bands. The exception was at Moderate danger, where forecasters and assessors accepted a greater range of AAI. That corresponds with the definition of Moderate, which the CAIC has associated with the broadest range of probable avalanche activity. Under-forecast days had a higher AAI, and over-forecast days had a lower AAI. Again, AAI played a critical role in the assessment process, reflected in the AAI values and critical factors.

4. CONCLUSIONS

Forecast verification is an integral part of the forecast process. Our hindsight assessments provide additional understanding of our public backcountry avalanche forecasts.

- The CAIC's forecast quality is high, with issued and assessed dangers matching most of the time. Our assessments found few major forecast misses.
- The T1 danger for the forecasts and assessments matched 84% of the time. Over- and under-forecasts were split equally, indicating little systematic bias.
- Avalanche activity was a key component during the assessments. The AAI of assessments varied less than the AAI of issued dangers. Some of the discrepancy can be explained by the forecast's emphasis of large, dangerous avalanches in old snow, while assessments could place more weight on less hazardous avalanches in recent snow.
- The assessments illustrated the importance of weather forecasts in the avalanche forecast process. On over-forecast days, the issued dangers were biased towards the upper end of forecast snow amounts to account for the potential of greater snowfall. If the forecast snow amounts did not verify, the assessments tended to a lower danger rating.
- Over-forecast danger ratings may be acceptable given the public safety application. While the CAIC does not want to be in a position of "crying wolf," preparing the recreating public for the most dangerous potential outcome may be beneficial and provide greater forecast Value than under-forecasting.

The assessment process indicated several areas the CAIC could improve operationally. One area is in weather forecasting and clearly explaining the forecast assumptions leading to the danger ratings. This might reduce the user's perception of a missed forecast if their observed danger is lower than is described in the forecast products.

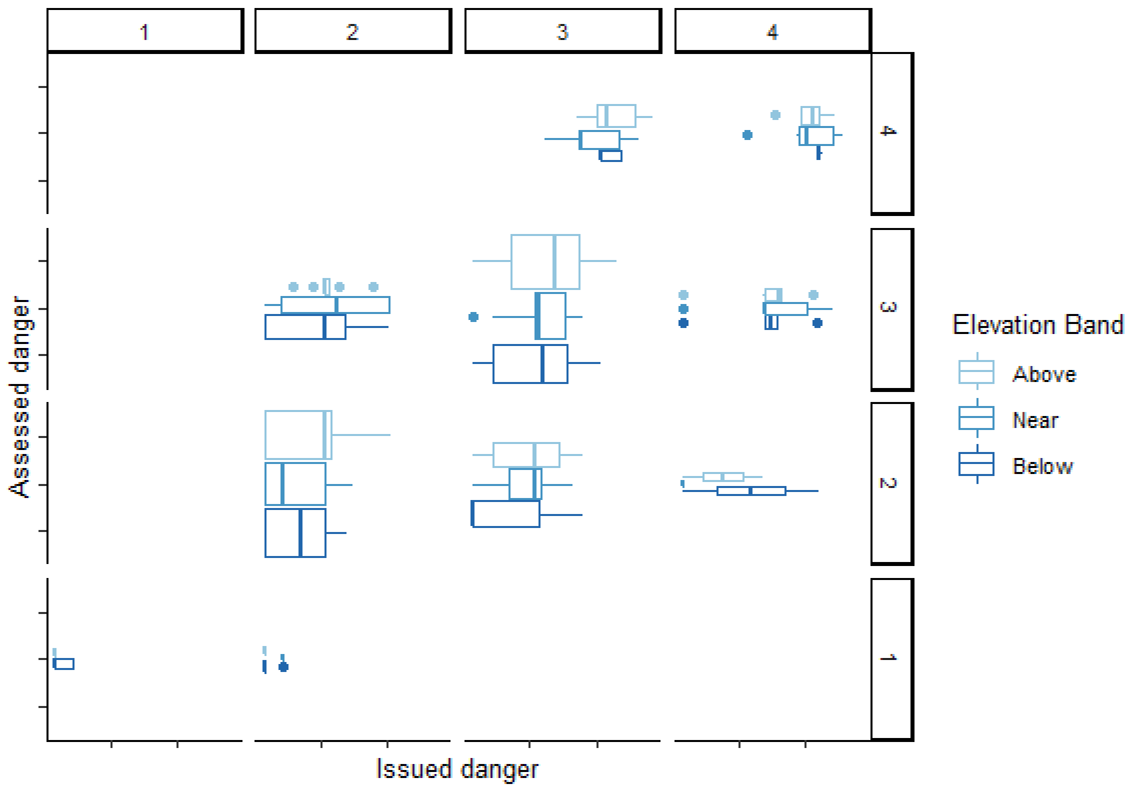

Figure 3. Box plots of log(AAI) by elevation band, presented similarly to the contingency tables. Columns are forecast danger ratings, while rows are assessed dangers. The center diagonal indicates correct forecasts based on assessments, about 84% of total assessments. AAI is exponential, so the log transform makes differences at lower values more visually apparent.

ACKNOWLEDGEMENTS

Many thanks to the Colorado Avalanche Information Center staff and everyone who reports and documents avalanches and snow conditions around Colorado. We benefited from conversations with colleagues worldwide, including Frank Techel, Scott Savage, and Jason Konigsberg. We used assessments conducted by Nick Barlow, Chris Bilbrey, Mike Cooperstein, Dylan Craaybeek, Jeff Davis, Ian Fowler, Jason Konigsberg, Brian Lazar, Spencer Logan, and Jeremy Yanko.

REFERENCES

Cagnati, A., M. Valt, G. Soratroi, J. Gavaldà, and Sellés, C.G. (1998). A field method for avalanche danger-level verification. Annals of Glaciology, 26, 343–346, https://doi.org/10.3189/1998aog26-1-343-346.

EAWS 2023, European avalanche danger scale. European Avalanche Warning Service. www.avalanches.org/downloads/#avalanche-danger-scale. Accessed August 2023

Ebert, P. A. and Milne, P. (2022). Methodological and conceptual challenges in rare and severe event forecast verification. Natural Hazards Earth System Sciences, 22(2), 539-557, doi.org/10.5194/nhess-22-539-2022.

Elder, K., and Armstrong, B. (1987). A quantitative approach for verifying avalanche hazard ratings. IAHS Publication,162, 593-603.

Föhn, P. M. B., and Schweizer, J. (1995). Verification of avalanche hazard with respect to avalanche forecasting, 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, 151–156.

Jamieson, B., C. Campbell, and Jones, A. (2007). Verification of Canadian avalanche bulletins including spatial and temporal scale effects. Cold Regions Science and Technology, 51, 204–213, doi.org/10.1016/j.coldregions.2007.03.012.

Lazar, B., Trautman, S., Cooperstein, M., Greene, E., and Birkeland, K. (2016). North American Avalanche Danger Scale: Are Public Backcountry Forecasters Applying it Consistently? Proceedings: International Snow Science Workshop 2016 Proceedings, Breckenridge, CO, USA. https://arc.lib.montana.edu/snow-science/item/2307

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K.L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., Plattner, G.-K.,Yohe, G. W., and Zwiers, F. W. (2010). Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). Available at www.ipcc.ch.

McClung, D. M., (2000). Predictions in avalanche forecasting. Annals of Glaciology, 31, 377–381, https://doi.org/10.3189/172756400781820507.

Mock, C. J. and Birkeland, K. W. (2000). Snow avalanche climatology of the western United States mountain ranges. Bulletin of the American Meteorological Society, 81, 2367-2392.

Murphy, A. H., (1993). What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. Wea. Forecasting, 8, 281–293, https://doi.org/10.1175/1520-0434(1993)008<0281:WIA GFA>2.0.CO;2.

Sharp, E.(2014). Avalanche Forecast Verification Through a Comparison of Local Nowcasts with Regional Forecasts. Proceedings of the International Snow Science Workshop, Banff, Canada, 475–480.

Schweizer, J., Kronholm, K., and Wiesinger, T. (2003). Verification of regional snowpack stability and avalanche danger. Cold Regions Science and Technology, 37(3), 277-288. doi.org/10.1016/S0165-232X(03)00070-3

Schweizer, J., Mitterer, C., Techel, F., Stoffel, A., and B. Reuter, B. (2020). On the relation between avalanche occurrence and avalanche danger level. The Cryosphere, 14, 737–750, doi.org/10.5194/tc-14-737-2020.

Statham, G., Haegeli, P., Birkeland, K., Greene, E., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., Kelly, J. (2010). The North American public avalanche danger scale. Proceedings of the 2010 international snow science workshop, Squaw Valley, CA, pp 117–123

Statham, G., Haegeli, P., Greene, E., Birkeland, K., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., and Kelly, J. (2018a). A conceptual model of avalanche hazard. Natural Hazards, 90(2), p 663-691.

Statham, G., Holeczi, S., and Shandro, B. (2018b). Consistency and Accuracy of Public Avalanche Forecasts in Western Canada. Proceedings of the 2018 International Snow Science Workshop, Innsbruck, Austria, 1491–1495.

Techel, F., (2020). On Consistency and Quality in Public Avalanche Forecasting – a Data-Driven Approach to Forecast Verification and to Refining Definitions of Avalanche Danger, PhD thesis, University of Zürich, Zurich, Switzerland, 227 pp.

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