

## AVALANCHE FORECAST VERIFICATION THROUGH A COMPARISON OF LOCAL NOWCASTS WITH REGIONAL FORECASTS

EIRIK SHARP<sup>1\*</sup>

<sup>1</sup>Canadian Avalanche Center, Revelstoke, BC Canada

**ABSTRACT:** This paper presents a methodology and a new dataset used by the Canadian Avalanche Center's Yukon Field Team in the winters of 2013 and 2014 to validate their forecasts of avalanche danger and primary avalanche problem. Using established techniques 77 local nowcasts were compared to operational and public regional forecasts over a variety of lead times. Preliminary results indicate that these forecasts accurately predicated observed conditions for 69% to 85% of cases. There was little difference in accuracy found between the forecasts of danger and avalanche problem and the length of forecasts lead times did not have a significant affect. Several sources of bias are identified that should be filtered in future studies and suggestions for new research are presented.

**KEYWORDS:** Yukon, public forecasting, forecast accuracy, forecast skill, temporal scale.

### 1. INTRODUCTION

"A forecast is like an experiment - given a set of conditions, you make a hypothesis that a certain outcome will occur. You wouldn't consider an experiment to be complete until you determined its outcome. In the same way, you shouldn't consider a forecast experiment to be complete until you find out whether the forecast was successful." – D.S. Wilks

In the winter of 2012 the Canadian Avalanche Center in collaboration with the Yukon Avalanche Association unveiled a new public avalanche forecast for the Yukon and Northern BC. The Yukon forecast region is characterized by an absence of the data streams that traditionally form the foundation of regional public avalanche forecasting in Canada. A pilot field program was tasked with developing systems to gather, synthesize and communicate field data from the region. This program faced a variety of operational challenges but also presented an exciting opportunity for experimentation and innovation. This paper details the methodology used by the CAC's Yukon field team over the winters of 2013 and 2014 to verify their forecasts.

As a pilot public safety program operating in a remote data-sparse area there were three primary

motivations to attempt a forecast verification study:

- 1) To monitor forecast quality by assessing both the degree of accuracy of the forecasts and how they may be improving over time.
- 2) To improve forecast quality by identifying areas of weakness.
- 3) To compare the quality of different forecast types by determining to what extent one forecast system gives better forecasts than another, and in what ways is that system better. In this case the field team wished to compare their ability to forecast avalanche danger with their ability to forecast avalanche character.

The Yukon forecast region is of medium size (over 2500km<sup>2</sup>) and field program consists of a team of two technicians gathering observations from the field between 3 to 6 days a week depending on weather and travel conditions. The Canadian Avalanche Center produces bi weekly public avalanche forecasts for the Yukon. In addition to these the Yukon field team produces daily operational forecasts. Both these forecasts are compiled in the Aval-X system and follow the same format. They both include a 3-day assessment of danger ratings for three elevation bands, a 1-day forecast of the three primary avalanche problems that are driving avalanche danger, and several short paragraphs of text.

Forecast validations aim to assess the quality of a forecast or the degree to which the forecast corresponds to what actually happened. Murphy (1993) identified that the key attributes of a quality forecast are accuracy and skill. Accuracy is the correspondence between the forecast and the truth (as represented by observations). Skill is a

---

\* *Corresponding author address:*

Eirik Sharp, Public Avalanche Warning Service,  
Canadian Avalanche Center,  
110 Mackenzie Ave, Revelstoke, BC V0E 2S0;  
tel: 250.837.6405;  
email: esharp@avalanche.ca

measure of the relative accuracy of the forecast against some reference forecast such as random chance.

An established way of assessing a regional avalanche forecast is to compare the forecast conditions level with local observations. Elder and Armstrong (1987) suggested that avalanche observations are the best indicator of snowpack instability. Meister (1995) identified that avalanche observations alone are not an applicable for verification at lower danger levels where natural avalanche may be infrequent. Under these conditions other avalanche observations must be applied to assess snowpack stability. Föhn and Schweizer, (1995) proposed that snowpack stability tests, combined with other observations, are best suited to verify the lower levels of avalanche danger. Jamieson et al.(2007) compared local drainage scale nowcasts to verify regional avalanche danger forecast ratings for a number of public forecast in Canada. These nowcasts were based on a synthesis of avalanche observations, formal snowpack tests and informal widespread observations of snow stability made by roaming field teams. This study adopted the same technique to verify forecasts of regional avalanche character and danger

## 2. METHOD

### 2.1 *Field Methods*

The size and data sparsity of the Yukon forecast region requires field workers to travel extensively by snowmobile and helicopter as well as on skis in order to access field sites. The field team collects daily avalanche, snowpack and weather observations from both field sites, study plots and automated weather stations distributed around the region. The team tries to cover as much terrain as possible in a field day and to balance high resolutions point observations such as snow pits with a wide spread low resolution observations that helps capture spatial variability. Observations taken included but were not be limited to: avalanche observations, snow profiles and formal and informal stability tests, widespread probing for snowpack structure and depth, observations of surface signs of instability, evolving snow conditions and weather.

At the end of each field day the team produced a local nowcast of avalanche conditions for the area they visited. The nowcast based on the synthesis of the teams observations from the day and consisted of an assessment of the above tree-line av-

alanche danger (low, moderate, considerable, high, or extreme) and an assessment of the avalanche problem (loose dry, storm slab, wind slab, persistent slab, deep slab, loose wet, wet slab, or cornice) that posed the greatest contribution to avalanche danger. This process was highly subjective, and the team was encouraged to evaluate and balance the strength and weight of the each observation rather than use a threshold sum approach in their evaluation. Each field team member was asked to arrive at a personal evaluation of the local conditions observed before meeting and deciding on the final nowcast rating through consensus. The team also recorded their confidence in their assessment.

Once the day's local nowcast had been finalized the team produced an operational and or public forecasts for the next 3 days. The above tree line avalanche danger and primary avalanche problem for each of the three days was recorded along with an assessment of the team's confidence in their forecast.

### 2.2 *Analytical Methods*

Only data records with a complete forecast-nowcast pair were selected for analysis. The distribution of values of the selected records was compared to total dataset by calculating the Spearman Ranked Correlation coefficient  $\rho$ .

Avalanche danger is a ranked value. For each danger forecast-nowcast pair the difference  $\Delta D$  can be calculated by subtracting the ranked nowcast value from the forecast value.

$$\Delta D = D_{Fx} - D_{Nx} \quad (1)$$

The difference indicates weather the forecast of avalanche danger is overly cautious or incautious.

Methods for verifying multi-category forecasts start with building a contingency table (Table 1) showing the frequency of forecasts and nowcasts. Values along the diagonal where the forecast matches the nowcast and  $\Delta D = 0$  are referred to as *hits* (Wiks, 1995).

	Forecast					
		Low	Mod	...	Ex	Total
Nowcast	Low	$n(O_1, F_1)$	$n(O_1, F_2)$	...	$n(O_1, F_5)$	$N(O_1)$
	Mod	$n(O_2, F_1)$	$n(O_2, F_2)$	...	$n(O_2, F_5)$	$N(O_2)$
	...	...	...	...	...	...
	Ex	$n(O_5, F_1)$	$n(O_5, F_2)$	...	$n(O_5, F_5)$	$N(O_5)$
	Total	$N(F_1)$	$N(F_2)$	...	$N(F_5)$	$N$

Table 1: Contingency table of avalanche danger

Forecast accuracy or hit rate is given by the ratio of number of forecasts that matches their associated nowcast to the number of total forecasts.

$$h = \frac{1}{N} \sum_{i=1}^5 n(O_i, F_i) \quad (2)$$

Forecast skill is assessed by comparing the forecast to random chance known using the Hanssen-Kuiper's Discriminant (Wilks, 1995).

$$HK = \frac{\frac{1}{N} \sum_{i=1}^5 n(F_i, O_i) - \frac{1}{N^2} \sum_{i=1}^5 N(F_i) N(O_i)}{1 - \frac{1}{N^2} \sum_{i=1}^5 [N(O_i)]^2} \quad (3)$$

Avalanche problem categories cannot be ranked in the same way as danger and so a calculation of  $\Delta D$  is not appropriate. However, avalanche problem forecast verification also requires a contingency table (Table 2).

	Forecast					
		Loose dry	Wind	...	Cornice	Total
Nowcast	Low	$n(O_1, F_1)$	$n(O_1, F_2)$	...	$n(O_1, F_8)$	$N(O_1)$
	Mod	$n(O_2, F_1)$	$n(O_2, F_2)$	...	$n(O_2, F_8)$	$N(O_2)$
	...	...	...	...	...	...
	Ex	$n(O_8, F_1)$	$n(O_8, F_2)$	...	$n(O_8, F_8)$	$N(O_8)$
	Total	$N(F_1)$	$N(F_2)$	...	$N(F_8)$	$N$

Table 2: Contingency table of avalanche problems

The categorical hit rate gives the accuracy of forecasting a particular category. This allows for the categories, which are most difficult to forecast to be identified.

$$h_i = \frac{1}{N_i} n(O_i, F_i) \quad (4)$$

Forecast accuracy and skill are calculated as before.

### 3. RESULTS

Field technicians collected data for 77 days over the course of the 2013 and 2104 winter seasons. The distribution of danger ratings is plotted in Fig 1. A total of 53-1 day lead, 39-2 day lead and 32-3 day lead forecasts were compared with their associated nowcasts. Although these datasets are small they were found to show a significant enough correlation to the overall distribution of data to be valuable (Table 1). The relative frequency of the forecast nowcast pairs for over the different time scales forecasts is plotted in Figures 1 and 2.

Forecast scale		1 Day		2 Day		3	
Danger Sample Size 77	Size	53		39		32	
	$\rho$	Obs	Fx	Obs	Fx	Obs	Fx
		0.99	0.95	1.0	0.98	0.99	0.90
Character Sample Size 49	Size	29		22		25	
	$\rho$	Obs	Fx	Obs	Fx	Obs	Fx
		0.99	0.99	1.0	1.0	0.99	0.94

Table 3: Correlation of between the selected forecast nowcast pairs for different time scales and the total data set.

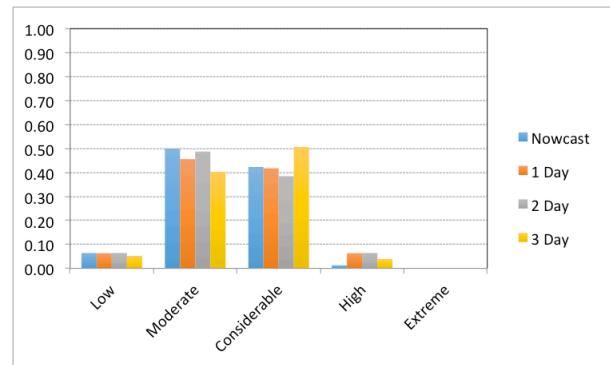


Figure1: Relative frequency of forecasts and nowcasts of danger ratings

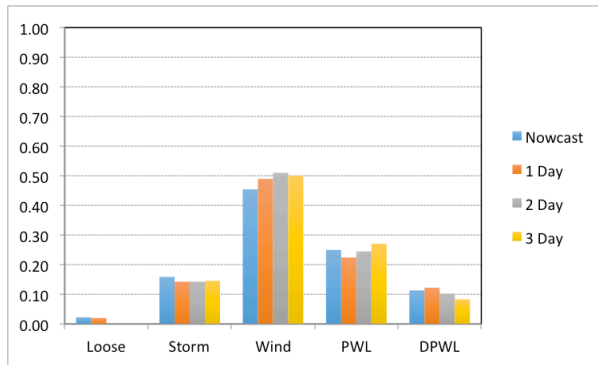


Figure 2: Relative frequency of forecasts and nowcast of avalanche problem

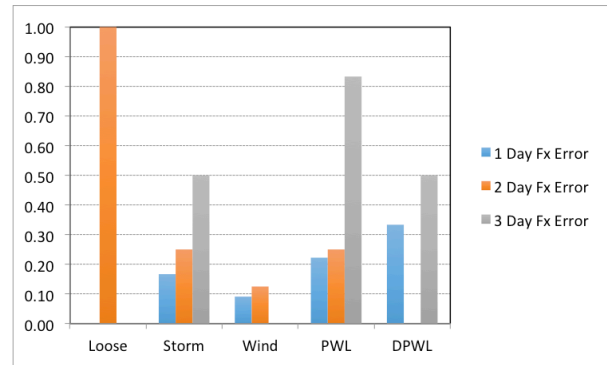


Figure 4: relative error frequency of forecast avalanche problem over different time scales.

The accuracy and skill scores of the forecasts are listed in Table 4. Skill score values greater than 0 indicated that the forecast performed better than random chance with 1 being perfect prediction.

Forecast Scale	Danger			Problem		
	1 Day	2 Day	3 Day	1 Day	2 Day	3 Day
h	81%	85%	69%	83%	73%	80%
HK	0.49	0.57	0.32	0.61	0.46	0.57

Table 4: summary of verification results for forecast of avalanche danger and avalanche problem over different time scales.

The error rate for the different time series of forecasts shows that forecasts of moderate or considerable avalanche danger were the most vulnerable to error (Figure 3). The sample size was too small to identify a clear distribution of errors for forecasts of avalanche problem (Figure 4).

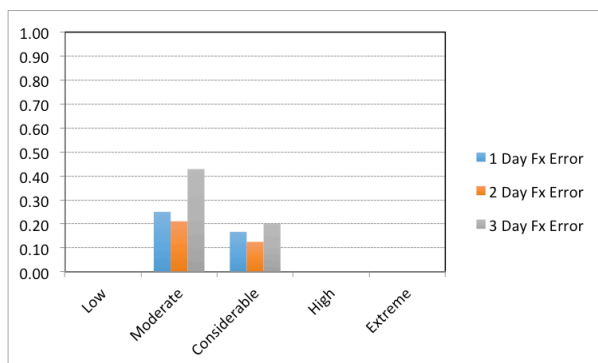


Figure 3: relative error frequency of forecast avalanche danger over different time scales.

The distribution of differences in avalanche danger forecast (Figure 5) indicates a tendency to over forecast avalanche danger. The relative distribution of categorical hit rates plotted in Fig 6 suggests a tendency to over forecast the influence of persistent weak layers especially on longer forecast periods (the high error associated with loose snow avalanches is likely due to there being only being one case in the data set).

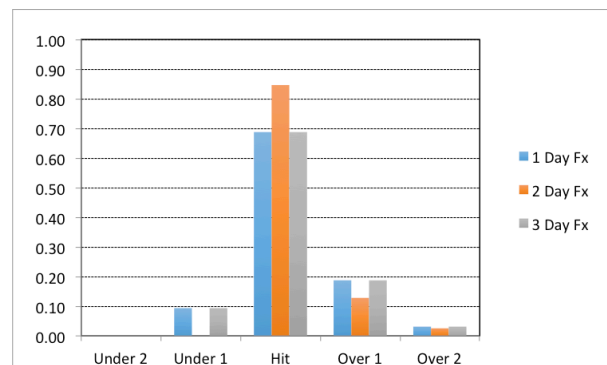


Figure 5: relative frequency of over or under forecast danger ratings over different time scales

## 4. DISCUSSION

### 4.1 Results

Although the sample size of this initial study is too small for to draw statistically significant conclusions, preliminary analyses offers some insight into the performance of the Yukon regional forecast. The results show that both avalanche danger and problem were forecast with similar levels of accuracy. The hit rate of both forecasts is high compared to other medium size forecast regions in Canada (Jamieson, 2007). This is very likely related to the distribution of data used in the analysis. The winters of 2013 and 2014 were characterized

by prolonged periods of stable weather and stagnant avalanche conditions. 74% of the forecasts predicted no change in avalanche danger and 80% predicted no change in avalanche problem over a given 3-day period. Avalanche danger was assessed to be either moderate or considerable in 92% of the nowcasts and wind slabs were observed to be the primary avalanche problem in 45% of the nowcasts. The tendency to over forecast avalanche danger is consistent with what has been observed in other regions.

For distributions of forecasts heavily weighted around a particular category or categories HK skill scores can often give a better indication of forecast quality than the hit rate. The HK skill scores for the Yukon's regional forecast are similar to those previously observed for other Canadian regional forecasts.

The initial results also suggest that time scales have little effect on the quality of the forecasts for the Yukon region. These results correspond with those for other regions of similar size. (Jamieson, 2007). However the lack of time scale effect may also be influenced by the consistency of avalanche conditions experienced over the course of the study.

#### 4.2 Biases

Several significant sources of bias were identified in this study. Efforts were made to reduce the influence of these biases but further work is required to analyze their effects.

The subjective nature of producing avalanche forecasts and nowcasts makes confirmation bias inevitable when forecasters endeavor to verify their own forecasts.

A tendency was identified for nowcast to be anchored on the forecast. The issue of anchoring bias has been encountered by previous studies. It is highly likely that prior knowledge of the forecast affects the nowcasts. However, Jamieson (2007) found that involvement in the production of the forecast did not have a significant effect on the forecast-nowcast hit rate when generating nowcasts based on roving observations.

Selection bias was also identified to have had a potential influence the results. Field sites were not randomly selected but were chosen to meet operational objectives. An effort to minimize the effect of selection bias was made by encourage field teams to cover as much terrain as possible each day and by basing nowcasts on a wide variety of

observations. The effectiveness of this strategy has not yet been evaluated.

#### 5. SUMMARY

The Yukon field team found that a forecast verification process was a useful and practical project. It gave the team confidence in their public avalanche forecasts and helped to identify and correct forecast weaknesses and repeating errors in a new and developing avalanche program. The preliminary results suggest that the CAC's Yukon regional forecast is of comparable quality to those produced for other regions in Canada.

Forecast verification results are more trustworthy when the quantity and quality of the verification data is high. Continuing to develop this dataset for the region over coming seasons will undoubtedly improve this verification. A larger dataset would also allow for further quantitative analysis of the biases implicit in this study.

#### 6. FURTHER STUDY

Any verification score must be regarded as a sample estimate of the "true" value for an infinitely large verification dataset. Uncertainty is associated with the score's value, especially in studies such as this where the sample size is small and the data is not independent. As this data set grows future studies should construct estimates of confidence intervals by using block-bootstrapping methods.

For the sake of simplicity this study focused on the verification of forecasts of only the primary avalanche problem. In reality the CAC's public avalanche forecasts often identify several different avalanche problems that contribute to avalanche danger in a bulletin. Methods that allow the analysis of forecasts with multiple avalanche problems would improve the verification.

Although the relationship between forecast quality and forecast confidence has yet to be analyzed anecdotal results suggest forecasters had greater confidence in their forecasts of avalanche problem than they did in their forecast of avalanche danger. This was not reflected by any significant difference in accuracy between the two types of forecast. Further analysis comparing forecast accuracy with forecaster confidence would be useful.

## 7. ACKNOWLEDGEMENTS:

Thanks to CAC Yukon field techs – Justin Abbiss, Scott Stewart, James Minifie, and Mat Holmes for collecting and recording data over the last two winters. Thanks also to CAC Public Avalanche Warning Service manager Karl Klassen and CAC Field program manager Ilya storm for support and to Dr James Floyer and Dr. Michael Smith for their guidance and mentorship.

## 8. REFERENCES:

Canadian Avalanche Association, 2002. Observation Guidelines and Recording Standards for Weather, Snowpack and Avalanches. Canadian Avalanche Association, Revelstoke, B.C.

Elder, K., Armstrong, B., 1987. A quantitative approach for verifying avalanche hazard ratings. In: Salm, B., Gubler, H. (Eds.), *Avalanche Formation, Movement and Effects*. International Association of Hydrological Sciences, Pub. No. 162, pp. 593–601.

Föhn, P.M.B., Schweizer, J., 1995. Verification of avalanche danger with respect to avalanche forecasting. In: Sivaridière, F. (Ed.), *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.

Jamieson, B., Campbell, C., & Jones, A. (2008). Verification of Canadian avalanche bulletins including spatial and temporal scale effects. *Cold Regions Science and Technology*, 51(2-3), 204–213.

Mason, S. J., & Mason, S. J. (2008). Understanding forecast verification statistics. *Meteorological Applications*, 15(1), 31–40.

Murphy, A. H. (1993). What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. <http://Dx.Doi.org/10.1175/1520->

Murphy, A. H., & Murphy, A. H. (1991). Forecast verification: Its Complexity and Dimensionality. *Monthly Weather Review*, 119(7), 1590–1601.

Schweizer, J., & Jamieson, J. B. (2007). A threshold sum approach to stability evaluation of manual snow profiles. *Cold Regions Science and Technology*, 47(1-2), 50–59.

Schweizer, J., Kronholm, K., Jamieson, J. B., & Birkeland, K. W. (2008). Review of spatial variability of snowpack properties and its importance for avalanche formation. *Cold Regions Science and Technology*, 51(2), 253–272.

Schweizer, J., Schweizer, J., Kronholm, K., Kronholm, K., Wiesinger, T., & Wiesinger, T. (2003). Verification of regional snowpack stability and avalanche danger. *Cold Regions Science and Technology*, 37(3), 277.

Wilks, D.S., 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego.