Evaluation of Avalanche Warning Strategies for Traffic Routes Using Signal Detection Theory

Christoph M. Rheinberger[1,*] and Jakob Rhyner[1]
[1] WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

ABSTRACT: Avalanche warning services (AWS) are operated to protect communities and traffic lines in avalanche prone regions of the Alps and other mountain ranges. In times of high avalanche danger, these services may decide to close a road or to evacuate a settlement. Their decisions are based on field observations, release statistics, forecasts issued by weather services and the experience of their operators. Due to the spatial variability in the snowpack and the insufficient understanding of avalanche triggering, these decisions are characterized by large uncertainty and the knowledge on which AWS have to base their safety-relevant decisions is incomplete. In this paper, we will show how signal detection theory (SDT) can be applied to make better use of the information that AWS have at hand. The proposed SDT-framework allows (i) the evaluation of past decisions on road closures and (ii) the improvement of the decision performance of AWS given their diagnostic ability and the disutility of decision failures. We will exemplify the use of this framework by evaluating the decision performance of two AWS and discuss the advantages of a formalized decision-making upon road closures.

KEYWORDS: Signal detection theory; Statistical performance measures; Diagnostic ability; Avalanche warning services (AWS).

1 INTRODUCTION

Technological improvements in forecasting and early warning systems have increased the use of active mitigation strategies in avalanche risk management (Bründl et al., 2004; Margreth et al., 2003; Rheinberger et al., 2009). Unlike structural control measures, active measures such as closing roads or evacuating settlements are only imposed in times of elevated avalanche danger. While they incur significantly less direct costs than structural measures, the effectiveness of these strategies depends greatly on the ability of avalanche warning services (henceforth AWS) to forecast hazardous situations and to close roads or evacuate endangered areas in a timely manner (Blattenberger and Fowles, 1995).

In their decisions AWS face the classical dilemma of choosing between accident avoidance and productivity losses as it is often costly to deny access to a ski resort, to interrupt commuting traffic from and to Alpine valleys, or to evacuate endangered settlements. Depending on the utilization of the road, interruption costs might be high (Blattenberger and Fowles, 1995) providing strong incentives to avoid false alarm closures. On the other hand, avalanches crossing an open road may cause fatal accidents that yield much higher costs, as the social cost per avalanche causality are estimated at €1.9-5.6 million (Rheinberger et al., 2009).

Both types of error—missed avalanche occurrences and false alarms—must be considered to prevent accidents in a cost-effective way. AWS typically feature flat hierarchies and decentralized decision-making, allowing their operators to quickly and adaptively react in hazardous situations and to deviate from routine. This enables the AWS to operate under the conditions of uncertainty inherent to avalanche forecasts, using decision rules that emerge from field experience and operational practice rather than preceding it (McClung, 2002).

Closing decisions are often made within hours and do not allow for a formal risk analysis. A fortiori, AWS operators should seek as much information on the likelihood of avalanche occurrences as they can obtain—notwithstanding that insufficient or inconsistent information and experiences may lead to ambiguous situations impeding their decision-making. Comparable to a physician who learns from past diagnoses and subsequent courses of disease, AWS operators may improve their closing performance by learning from past decisions and outcomes.

In this paper, we illustrate (i) how signal detection theory (Swets, 1996) can be applied to evaluate the performance of AWS and (ii) how AWS may improve their decisions based on the results of this evaluation. We first define the decision problem that AWS are faced with. Then, we formulate an framework to evaluate the decision performance of AWS, which could be adapted to decisions about other road hazards as well. To exemplify the use of this framework, we evaluate the decision performance of two Swiss AWS.

Corresponding author address: Christoph Rheinberger, WSL Institute for Snow and Avalanche Research SLF, Davos Switzerland;
tel: +41 81 417 0357; fax: +41 81 417 0110;
email: rheinberger@slf.ch

2 DECISION PROBLEM OF AWS

Swiss AWS rely on a decision system composed of three coupled components. Early warnings issued by the national avalanche forecasting service indicate meteorological conditions that substantially increase the probability of avalanche releases. These warnings are issued to local AWS whenever the probability for new snow depths of 1m or more over the next three days exceeds 40%. Triggered by early warnings, local AWS align the meteorological forecasts with field observations to obtain a regional forecast. Alerted by regional forecasts, the local AWS may assess the danger at particular avalanche paths. Based on these site-specific assessments, the operators then face the decision whether or not to close or re-open a road.

The decision tree in Fig. 1 shows the options the operators have: they can either close the road ($\delta = 1$) or keep it open ($\delta = 0$). Once this decision is made, an avalanche may occur ($\theta = 1$) or may not occur ($\theta = 0$). This leads to four potential outcomes, each described by a disutility vector $D(\delta,\theta)$.

However, due to a lack of understanding of the complex physical processes underlying snow avalanche formation (Schweizer et al., 2003), expert knowledge and field experience still play a key role within avalanche forecasting and warning (McClung, 2002). The local variability in the snowpack makes accurate forecasting a challenging task, which has led many AWS operators to adopt a “safety-first” strategy, although such precautionary behavior may lead to overprotective decisions at the extra costs of unnecessary road closures.

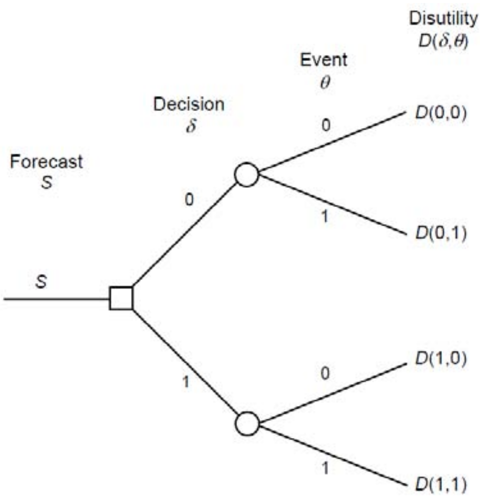

Figure 1. Decision tree for AWS.

The four potential outcomes of the decision problem are multidimensional and include economic, social and behavioral components. In the trivial case, the operators decide to keep the road open and no avalanche occurs ($\delta = 0, \theta = 0$). Given that all four outcomes require the same cost to operate the AWS, this *correct rejection* does not entail any additional cost and its outcome-specific disutility function is assumed to be neutral: $D(0,0) = 0$.

In uncertain situations such as after periods of heavy snowfall on peak traffic days, the operators may decide to close the road, but eventually no avalanche occurs ($\delta = 1, \theta = 0$). This *false alarm* causes direct costs c for the intervention and indirect economic losses i due to business interruption. Further, if the operators act overcautiously producing many false alarms, this may entail a loss of confidence into the forecasting system (Williams, 1980) and into the forecasting authorities, which we here denote by η. The disutility of a false alarm is written as: $D(1,0) = f_{10}(c,i,\eta)$.

Evidently, closing the road should coincide with avalanche occurrences as often as possible ($\delta = 1, \theta = 1$) in order to approve the site-specific forecast. Though effective, such a *hit* still entails direct costs c for the intervention as well as indirect economic losses i due to business interruption, and the avalanche occurrences may cause additional costs by damaging infrastructure denoted by s. On the other hand, successful closures stabilize confidence into the forecasts and may foster their acceptance among the affected road users. The disutility function of hits can be written as: $D(1,1) = f_{11}(c,i,s)$.

The worst outcome can occur when one or more avalanches cross the open road ($\delta = 0, \theta = 1$). Such a *miss* may lead to fatal accidents involving casualties, injuries, and emotional distortions as well as to a loss of confidence into the forecasts. Therefore, the disutility of misses is notated by the most complex disutility function: $D(1,0) = f_{01}(e,m,\omega,\eta,s)$, where e denotes the exposure of cars to the occurring avalanche, m denotes the mortality rate of occupants given a car gets hit, ω denotes the injury rate of occupants, η is the loss in confidence into the road-closing policy and s is the potential damage of infrastructure.

By this definition, the disutility of a miss $D(0,1)$ depends on the site-specific probability of an avalanche accident and the consequences thereof. As the probability and the consequences of an accident are not predictable with certainty, we may model them by defining random variables E for the exposure and M for the mortality rate conditional on the exposure. For casualties, the expected disutility of a miss becomes:

$$\mathrm{E}[D(0,1)] = \int_E \int_M P(e) \times P(m|e) \times f_{01}(e,m,\eta,s)\, \mathrm{d}e\, \mathrm{d}m, \quad (1)$$

where $P(e)$ and $P(m|e)$ denote the probability density functions of the random variables. Eq. (1) illustrates that the outcome-specific disutility of a miss does not necessarily have to be larger than that of a false alarm or a hit as—in many cases—a miss would not result in an accident.

For a particular road, the outcome-specific disutility can be determined via monetary estimates for the disutility functions. The operators of the responsible AWS face the classical problem of decision-making under risk, requiring them to minimize the expected disutility that comes along with the decision δ:

$$\min \mathrm{E}[D(\delta)] = \sum_{\delta}\sum_{\theta} P(\theta) \times P(\delta|\theta) \times D(\delta,\theta), \tag{2}$$

where $D(\delta,\theta)$ denotes the disutility of each decision outcome, $P(\theta)$ is the probability of an avalanche (non-)occurrence, and $P(\delta|\theta)$ is the conditional probability to decide appropriately given the (non-)occurrence of avalanches.

It is not known in advance whether and when an avalanche will occur, but $P(\theta)$ at a particular avalanche path can be approximated by the frequency of past avalanches. $P(\delta|\theta)$ can be deduced from past decisions and their success or failure. In section 3, we build on earlier research from meteorology (Brooks, 2004; Harvey et al., 1992) to develop an SDT-framework that allows evaluating this decision performance.

3 EVALUATION FRAMEWORK

SDT provides a theoretical framework to deal with any forecasting process (Swets, 1996). With respect to road-closing policies, the operators of an AWS must decide whether the site-specific forecast indicates to close a road or not. This decision leads to the four outcomes discussed above, whose relative frequencies can be organized in a 2 × 2 contingency table.

Based on these relative frequencies, three statistical performance measures have been established that can be used to evaluate road-closing decisions (Wilks, 2006). The false alarm ratio (FAR) measures the diagnostic ability:

$$\text{FAR} = 1 - P(\theta = 1|\delta = 1) = \frac{\pi_{10}}{\pi_{10} + \pi_{11}}, \tag{3}$$

where FAR = 0 means that every road closure is followed by one or more avalanche occurrences implying perfect diagnosis. The probability of detection (POD) characterizes the reliability of the AWS:

$$\text{POD} = P(\delta = 1|\theta = 1) = \frac{\pi_{11}}{\pi_{01} + \pi_{11}}, \tag{4}$$

where POD = 1 means that avalanche occurrences are never missed.

A third measure that states how many safe situations are misinterpreted as being dangerous leading to false alarms is the probability of false detection (POFD):

$$\text{POFD} = P(\delta = 1|\theta = 0) = \frac{\pi_{10}}{\pi_{10} + \pi_{00}}, \tag{5}$$

where POFD = 0 means that no safe situation is wrongly interpreted as dangerous.

3.2 Decision thresholds

In a particular decision, the FAR, POD and POFD depend on how much weight is given to the observed evidence such as the occurrence of avalanches on neighboring paths, the overall snow depths, and the new snow depth. The weight that the operators attach to the observed evidence can be conceived as a decision threshold t (Brooks, 2004). Whenever the weight of evidence is above t, the road should be closed.

The probability that the evidence for a road closure exceeds the threshold given that one or more avalanches occurred is given by $P(\text{Evidence}_\delta \geq t \mid \theta = 1)$, and the probability that the evidence for a road closure is below the threshold given that no avalanche occurred is given by $P(\text{Evidence}_\delta < t \mid \theta = 0)$. SDT assumes that the operators are able to establish the decision threshold t so that it reliably distinguishes an event *signal* from the non-event *noise*.

Since signal and noise may interfere, evidence for closing a road is often ambiguous. To statistically separate them, the SDT-framework assumes that signal and noise are normally distributed variables (Abdi, 2007). For ease of computation, let us assume that the noise is a standard normally distributed random variable $X(\tau) \sim N(0,1)$ and the signal is a normally distributed random variable $Y(\tau) \sim N(\mu_Y,\sigma_Y)$. POFD and POD for any decision threshold t may then be described by the complementary error functions of $X(\tau)$ and $Y(\tau)$:

$$\text{POFD} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp(-\tau^2/2)\,\mathrm{d}\tau, \text{ and} \tag{6}$$

$$\text{POD} = 1 - \frac{1}{\sigma_Y\sqrt{2\pi}} \int_{-\infty}^{t} \exp[-(\tau-\mu_Y)^2/2\sigma_Y{}^2]\,\mathrm{d}\tau. \tag{7}$$

The difference between the means of both distributions is called the distance d' and reflects the decision maker's ability to separate between signal and noise (Abdi 2007). Since the mean of the noise distribution is normalized to zero, d' equals the mean of the signal distribution. Based on d' the ideal observer, which is the most popular standard for the decision threshold t, can be derived as $t^* = d'/2$. It jointly minimizes the probability of misses and false alarms. However, since in most cases $D(1,0) \neq D(0,1)$, one may use what we call the utility maximizer's threshold: $t_D = d'/2 \times [D(1,0)/D(0,1)]$. Depending on the ratio between the disutility of misses and false alarms, this threshold moves toward/away from the mean of the noise distribution.

3.3 ROC Curves

FAR, POD and POFD can be estimated from data on past road closures. However, this static perspective on the decision performance may not be sufficient, as it does not allow the assessment of changes over time in the diagnostic ability of an AWS. Alternatively, receiver operating characteristics (ROC) graphs might be used (Wilks, 2006) to depict the locus of all pairs of values {POFD, POD}, illustrating the ideal observer and other possible thresholds (Fig. 2).

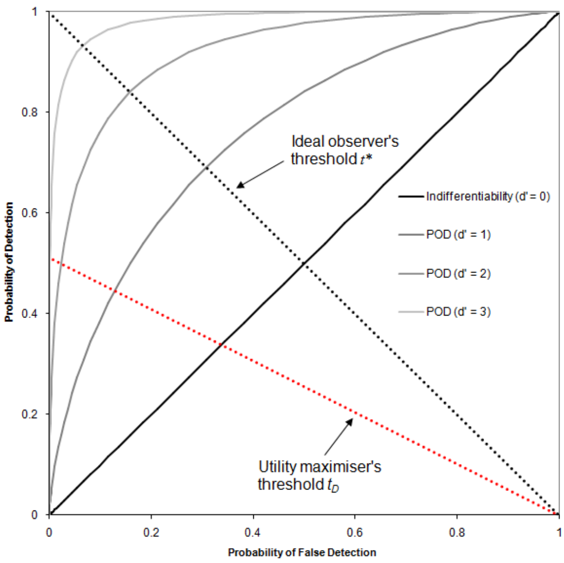

Figure 2. ROC graphs for different d' values. The red line indicates the ideal observer t^*, the black line exemplifies the utility maximizer's threshold t_D assuming that $\{D(1,0)/D(0,1)\} = 1/2$.

From Fig. 2, it is obvious that if d' increases the performance of the decision maker increases as well. As d' is not known *a priori*, it has to be estimated from the observed data by $d' = \Phi^{-1}(\text{POD}) - \Phi^{-1}(\text{POFD})$, where Φ^{-1} denotes the inverse normal distribution. POD and POFD are estimated from Eqs. (4) and (5).

4 CASE STUDY APPLICATION

We now apply the outlined SDT-framework to evaluate the road closing decisions of two Swiss AWS, highlighting the options that they would have by changing their decision strategy. The first case (AWS#1) draws on observations from 1991/92 to 2004/05 of an AWS, monitoring a single road with multiple avalanche paths in Eastern Switzerland, the second case (AWS#2) draws on observations from 1990/91 to 2006/07 of an AWS in Western Switzerland, monitoring several roads with single or multiple paths.

As mentioned by Brooks (2004), the determination of non-events is a thorny task. We approximate it by the relative frequencies of the four outcomes based on the mean number of days per winter on which the Swiss bulletin forecasts the avalanche danger to reach or exceed the third level on the European avalanche danger scale: n = 54 days (SLF, 2007). On these days, we expect considerable avalanche danger as the snowpack is moderately to weakly bonded on many steep slopes. Hence, we restrict the event space making stratification between trivial and difficult decision situations (Murphy, 1995).

Multiplication of these danger days by the period of observation gives the total number of avalanche-prone days (APD) within the samples. The observed hits, misses, and false alarms were then divided by the APD to obtain their relative frequencies. Correct rejections could not be directly observed, but were assumed to hold the share of the APD on which none of the described events was observed. This procedure allows compiling the POD, FAR, and POFD as well as the distance d' between signal and noise to outline the differences in the decision performance of the two AWS.

AWS #2 had a distinctly higher POD, a distinctly lower FAR and a somewhat lower POFD than AWS #1 (Table 1). The FAR shows that the operators of AWS #2 discriminated better between noise and signal. This difference might be explained by the fact that the monitored road system in Western Switzerland consists of multiple roads with independent avalanche paths, whereas AWS #1 watches over a single road with multiple interdependent avalanche paths.

Closing decision	AWS#1: Avalanche occurrences θ		AWS#2: Avalanche occurrences θ	
δ	1	0	1	0
1	0.04	0.15	0.09	0.10
0	0.05	0.76	0.01	0.80

Table 1: 2 × 2 contingency table of the decision outcomes for the analyzed AWS.

However, even AWS#1's probability to correctly issue a road closure was about three times higher than to give a false alarm, implying that their warning and forecast system has good ability to discriminate between situations of low, medium, and high avalanche danger. Moreover, if we assume the disutility of a miss larger than that of a false alarm, decision behavior under risk suggests that neutral decisions, such as by the ideal observer's threshold t^*, are no longer desirable (Brooks, 2004). In that case, a higher POD at the costs of a higher FAR might be well accepted by the decision maker.

Based on the estimated distance *d'*, we construct a receiver operating characteristic (ROC) graph (Wilks, 2006), wherein we mark the actual position of both services on their respective ROC curve (Fig. 3). Obviously, both AWS have so far given more weight to the prevention of misses as their actual position is below the ideal observer's threshold *t**.

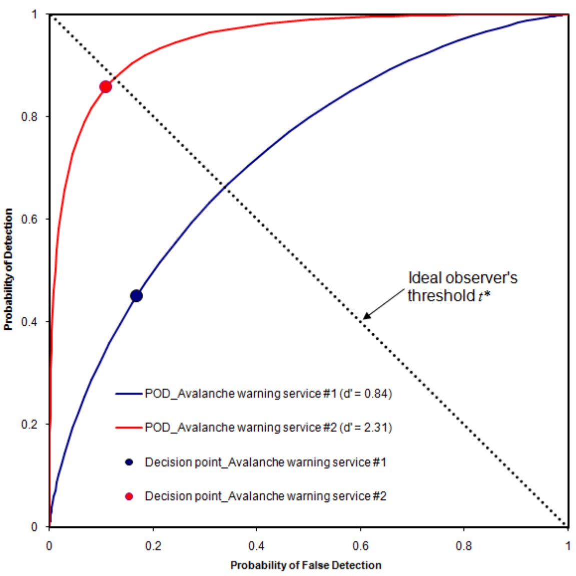

Figure 3. ROC graphs for the AWS under analysis. Their actual decision points are marked on the curves.

5 CONCLUSIONS

A major task of AWS is to maintain high levels of safety and accessibility on Alpine roads. Balancing safety and productivity losses caused by road closures is thus an essential objective for the design and performance of AWS. So far, AWS have been operated mainly based on the experiences of their staff. In the preceding sections we have presented a generic SDT-framework, which might help the operators of these AWS to evaluate their statistical decision performance in road closures. As this performance varies with the complexity of the decision task, the proposed framework can also be used to compare the diagnostic ability of different AWS under varying conditions of the monitored road systems.

This diagnostic ability is—at least in the short run—inert and improvements in the POD come at the cost of increasing the FAR. The decision performance analysis of two exemplary AWS has shown that there is scope to improve the discrimination between noise and signals of avalanche danger. Learning from past decisions is one key aspect for such an improvement and the presented framework provides the tools to analyze data of past road-closing decisions, helping AWS to exploit the data they have at hand in a formalized way.

6 REFERENCES

Abdi, H., 2007. Signal Detection Theory (SDT). In: N.J. Salkind (Editor), Encyclopaedia of Measurement and Statistics. Sage, Thousand Oaks, CA, pp. 886-889.

Blattenberger, G. and Fowles, R., 1995. Road Closure to Mitigate Avalanche Danger – a Case-Study for Little Cottonwood Canyon. Int. J. Forecasting, 11(1): 159-174.

Brooks, H.E., 2004. Tornado-warning performance in the past and future - A perspective from signal detection theory. Bull. Am. Meteorol. Soc., 85(6): 837-843.

Bründl, M. et al., 2004. IFKIS – a basis for managing avalanche risk in settlements and on roads in Switzerland. Nat. Hazards Earth Syst. Sci., 4(2): 257-262.

Harvey, L.O., Hammond, K.R., Lusk, C.M. and Mross, E.F., 1992. The Application of Signal Detection Theory to Weather Forecasting Behavior. Mon. Weather Rev., 120(5): 863-883.

Margreth, S., Stoffel, L. and Wilhelm, C., 2003. Winter opening of high alpine pass roads – analysis and case studies from the Swiss Alps. Cold Reg. Sci. Technol., 37(3): 467-482.

McClung, D.M., 2002. The elements of applied avalanche forecasting - Part I: The human issues. Nat. Hazards, 26(2): 111-129.

Murphy, A.H., 1995. A coherent method of stratification within a general framework for forecast verification. Mon. Weather Rev., 123: 1582-1588.

Rheinberger, C.M., Bründl, M. and Rhyner, J., 2009. Dealing with the White Death: Avalanche Risk Management for Traffic Routes. Risk Anal., 29(1): 76-94.

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

SLF, 2007. Avalanche Bulletins and other products of the Swiss Federal Institute for Snow and Avalanche Research SLF. SFL, Davos.

Swets, J.A., 1996. Signal detection theory and ROC analysis in psychology and diagnostics. Erlbaum Associates, Mahwah, NJ.

Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences. Elsevier, Amsterdam.

Williams, K., 1980. Credibility of Avalanche Warnings. J. Glaciol., 26(94): 93-96.