Bayesian networks to quantify the reliability of a debris flow alarm system

M. Sättele & M. Bründl
WSL Institute for Snow and Avalanche Research SLF, Avalanche Dynamics and Risk Management, Davos Platz, Switzerland

D. Straub
TUM Technische Universität München, Engineering Risk Analysis Group, München, Germany

ABSTRACT: Warning and alarm systems are part of an integrative approach to risk management for natural hazards that supplement protective measures such as rock fall nets, dams and galleries. To integrate warning and alarms systems as standard measures in a risk management approach, their reliability should be quantified. In this paper, selected methods are applied to quantify the reliability of an active threshold-based debris flow alarm system. The reliability is defined as the ability of the system to detect dangerous debris flow events, to issue alarms in a timely manner and to avoid false alarms. Bayesian networks are applied to probabilistically model the considered alarm system and to calculate its overall reliability. The final system reliability is expressed in terms of the receiver operator characteristics, which allow the identification of the optimal trade-off between the probability of detection and the probability of false alarms.

1 INTRODUCTION

1.1 *Warning and alarm systems in an integrative risk management approach for natural hazards*

An integrative approach to risk management incorporates risk mitigation measures to reduce damage caused by natural hazards. In recent years, warning and alarm systems (WASs) have been increasingly applied as cost-efficient risk mitigation measures with little environmental impact (Intrieri et al., 2013). WASs generate information or issue an alert before a hazard event causes loss of property and life (UNEP, 2012). By generating these alerts, WASs mainly mitigate the overall risk by reducing the probability of presence of endangered persons. The vulnerability or the number of mobile objects can only be reduced if sufficient lead time is offered and appropriate countermeasures are taken (Medina-Cetina & Nadim, 2008, Bründl et al., 2009). To incorporate WASs as standard measures into an integrative risk management approach and to make them comparable to well proven measures such as rock fall nets, dams and galleries, their reliability should be quantified.

1.2 *A system classification to identify factors relevant to reliability*

To identify factors that influence the reliability of the currently operational WASs in a structured manner, Sättele et al. (2012) classify WASs into three main categories: i) threshold systems, ii) expert systems and iii) model-based expert systems. They apply the classification to 52 WASs in Switzerland and identify typical system characteristics for each system class. As a result, the classification allows a structured identification of factors that influence the reliability of each defined system class.

The reliability of a threshold system strongly depends on the definition of an appropriate alarm threshold, on the availability of well-designed and redundant sensor units and on the functionality of the alarm system to start and stop transmission of the warning. The alarm threshold of WASs is a predefined minimum value of monitored parameters. If this value is exceeded, an alarm is automatically initiated. The threshold is an essential variable to characterize and distinguish events in an early warning process (Felgentreff & Glade, 2008).

In this paper, the reliability of an active threshold system, the debris flow alarm system at Illgraben, is quantified. Suitable reliability methods, which address the factors influencing the reliability of a threshold system, are identified, applied and evaluated. This case study is the first step towards the development of a generic method to quantify the reliability of WASs, which shall ultimately help incorporating WASs as standard measures into an integrative risk management approach.

2 RELIABILITY METHODS

2.1 *System reliability analysis*

In many industry sectors, tailored standards and methods exist to assess system reliabilities. Reliability methods such as Failure Mode Effects and Criticality Analysis (FMECA) and Fault Tree Analysis (FTA) are applied in production industries to evaluate the reliability of the system designs in different life cycle phases. Thereby, it is common to express the reliability of system components in failure rates (Blanchard & Fabrycky, 2011). In most classical reliability methods, dependencies between the system components are not considered (Pagès & Gondran, 1986). Specially designed methods for quantification of the reliability of WASs for natural hazards could not be identified. An initial approach to investigate the reliability of the Swiss avalanche warning system was presented by Bründl & Heil (2011). They use FTA to identify the most critical system components, but conclude that the method is not sufficient to cover the total complexity of the system. Natural hazard events are complex, often site-specific, rare and random in nature. WASs that are currently operated for a variety of natural hazard processes are mainly installed as prototypes. To quantify the overall system reliability, methods used in the area of civil engineering appear to be suitable. They are designed for the assessment of unique, complex systems under uncertainties (Bensi et al., 2012). The reliability of civil engineering systems are assessed through combining probabilistic models of system parameters with physical or logical models representing system failures (Stewart & Melchers, 1997).

2.2 *Bayesian Networks to quantify the reliability of warning and alarm systems for natural hazards*

Bayesian Networks (BNs) are suitable methods for probabilistic quantification of the reliability of WASs for natural hazards. Aquilera et al. (2011) summarize suitable applications of BNs for environmental modeling and Straub (2005) states reasons why BNs have a large potential for assessing natural hazard risks. BN allow the incorporation of expert knowledge, deal with rare data and are based on an intuitive modeling approach (Friis-Hansen, 2001). A BN is a graphical model that allows a quantitative system analysis by combining sub-system reliabilities. The nodes of a BN represent random variables to express the uncertain state of a component. The arcs represent the causal probabilistic dependencies between the nodes (Pearl 1988, Jensen & Nielsen 2007). Nodes have a child-parent relationship and the probabilities of the states of a child node are expressed in conditional probability tables (CPTs) (Jensen & Nielsen, 2007). For this case study, the free software GeNIe is used to calculate the reliability of the Illgraben debris flow alarm system (DSL, 2013).

2.3 *Receiver Operator Characteristics to express the system performance*

Receiver Operator Characteristics (ROCs) are combined with the BN to represent the overall system performance as the ratio between the probability of detection POD (true positives) and the probability of false alarms PFA (false positives) (Figure 1). ROC curves have their roots in signal detection theory, but have seen widespread application, e.g. to represent the accuracy of diagnostic tests in medicine (Fawcett, 2006). ROC curves are a simple tool for the definition of optimal alarm thresholds (England, 1988, Schoefs et al., 2009). An optimal threshold allows the detection of all events and keeps false alarms to a minimum (Intrieri et al., 2013). In this case study, ROC curves are modeled to investigate the influence of system alarm thresholds on the system reliability.
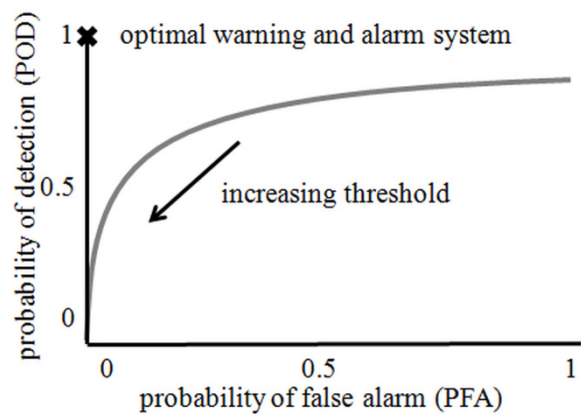

Figure 1. Typical Receiver Operator Characteristic (ROC) curve to model the relationship between the probability of detection and the probability of false alarm.

3 THE ILLGRABEN DEBRIS FLOW ALARM SYSTEM

3.1 *The Illgraben: An active debris flow catchment*

The following analysis aims to quantify the reliability of a threshold system, which is installed in Switzerland to detect dangerous debris flows and floods in the Illgraben catchment. A debris flow is a fast-flowing mixture of water and solid particles which typically consists of surges. The Illgraben catchment ranges in elevation from 610 m a.s.l. to 2716 m a.s.l. and half of the catchment area (~ 4 km^2) is covered by bedrock and debris deposits. Due to the geological conditions there is a remarkably high occurrence rate of debris flows (Badoux et al., 2009). In 2000, a monitoring system was installed by the Swiss Federal Institute for Forest, Snow and Landscape Research WSL, which was further developed into an alarm system in 2006. Since 2008 a detailed data base has been built up. It contains continuously measured sensor data that are included in this reliability analysis.

3.2 Design of the alarm system

Two sensor units are located close to the release area to detect events in real-time. Sensor unit 1 (SU1) is located in the upper catchment, where geophone 1 (G1) continuously monitors ground vibrations. Sensor unit 2 (SU2) is located 200 m below, where two geophones (G2 / G3) and two radar devices (R1 / R2) measure the ground vibrations and the flow depth. The sensors in each unit are controlled by a logger and if predefined threshold values are exceeded, an alarm call is automatically issued via modem and transmitted to the communication unit in the valley. The power at these remote locations is supplied via solar panels and batteries. The communication unit in the valley forwards incoming alarm calls via T-Box and SMS-Butler to three alarm units. They are located close to three crossings of the streambed in the catchment. Each alarm unit is equipped with a red light and an audible alarm. In addition, responsible authorities and system operators receive alarm messages via SMS and E-mail.

The alarm time for endangered persons is in the range of 5-15 min and is determined by the distance between the sensors and the uppermost crossing in the channel. At present, the system aims to detect all events that represent a hazard, even for persons crossing the catchment (Graf et al., 2009). For the case study, events are classified according to the flow depth increase, measured by R1 and R2. At this location, small events (A) lead to a flow depth increases of up to 0.6 m. Further down the catchment, close to the crossing, they remain entirely within the channel and represent a hazard only for persons situated in the channel. Medium-sized events (B) have flow depth increases of up to 1.5 m and imply a limited overbank flow. Large flood events (C) are characterized through flow depth increases of more than 1.5 m and are characterized through small amounts of debris. Debris flows events (D) contain a high debris percentage. Since 2000, D events varied in density and reached flow depths of between 1 m and 8 m, but never entered populated areas. According to historical documentation, the last serious overbank flow happened in 1961 (Badoux et al., 2009).

4 BAYESIAN NETWORKS TO QUANTIFY THE RELIABILITY OF THE ILLGRABEN SYSTEM

4.1 Three levels of the Bayesian Network

In this case study, BNs are applied to probabilistically model the alarm system and quantify its reliability. The overall system reliability refers to the ability of the system to fulfill its designated performance, which is to issue an alarm to protect endangered objects close to the channel of an ongoing event in a timely manner and to avoid false alarms. The BN developed for Illgraben consists of three main levels.

Level one represents the main functionalities and information flow through the BN. Level two models the technical components and their dependencies. Level three focuses on the interpretation of the measured data and deals with the influence of the system threshold on the overall system reliability.

4.2 Main functionalities and information flow

The basic structure of the BN reflects the system's main functionalities and information flow. Main system functionalities are represented as chance nodes and the arcs show the information flow between these functionalities (Figure 2).
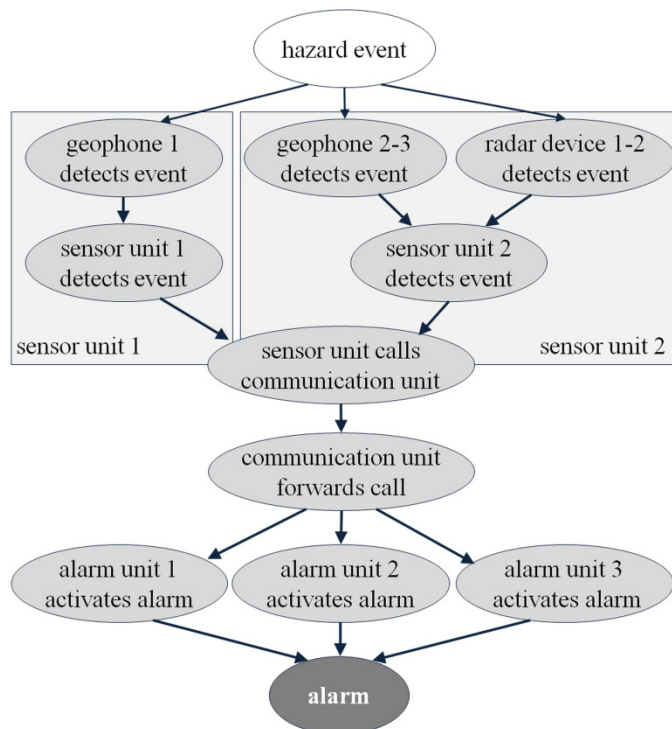

Figure 2. Main functionalities (chance nodes) and information flow (arcs) through the BN.

The CPTs of nodes representing the main functions allow the definition of terms and conditions under which information is forwarded to the child nodes. Theses nodes are deterministic, i.e. the probability values in their CPTs are 1 or 0. Due to the redundant sensor design, the state of the node "sensor unit calls communication unit" is given by an OR condition (Figure 3). The state is set to "yes" if either SU1 or SU2 indicates event detection or issues a false alarm. Therefore, the state of either the parent node "sensor unit 1 detects event" OR the parent node "sensor unit 2 detects event" is "yes".

sensor unit 1 detects event		yes		no	
sensor unit 2 detects event		yes	no	yes	no
sensor unit calls	yes	1	1	1	0
communication unit	no	0	0	0	1

Figure 3. CPT of the node "sensor unit calls communication unit": Main functionality is based on OR conditions.

In contrast, the state of the root node "alarm" is defined through an AND condition (Figure 4). The root node only indicates that an event is detected or a false alarm issued if the states of all three child nodes "alarm unit 1 / 2 / 3 activates alarm" are "yes".

alarm unit 1 activates call	yes				no			
alarm unit 2 activates call	yes		no		no		yes	
alarm unit 3 activates call	yes	no	yes	no	yes	no	yes	no
alarm yes	1	0	0	0	0	0	0	0
alarm no	0	1	1	1	1	1	1	1

Figure 4. CPT of the node "alarm": Main functionality is based on AND conditions.

4.3 Technical components and dependencies

The states of chance nodes, which represent main functionalities, are moreover directly influenced by the sub-reliability of individual and dependent technical system components (TSCs). The states of these TSCs are expressed through survival or failure probabilities. The functionality "sensor unit calls communication unit" is only guaranteed if the modem is functioning and is supplied with power by the battery and the solar panel (Figure 5). The CPT of "sensor unit calls communication unit" is extended to the one in Figure 3 to include the influence of TSC failures (Figure 6).
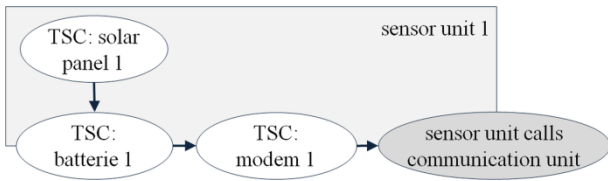

Figure 5. Technical system components influence the functionality of the node "sensor unit calls communication unit".

TSC: modem 1	survival				failure			
TSC: modem 2	survival		failure		survival		failure	
sensor unit 1 detects event	yes	no	yes	no	yes	no	yes	no
sensor unit 2 detects event	yes no	yes no	yes no	yes no	yes no	yes no	yes no	yes no
sensor unit calls yes	1 1	1 0	1 1	0 0	1 0	1 0	0 0	0 0
communication unit no	0 0	0 1	0 0	1 1	0 1	0 1	1 1	1 1

Figure 6. CPT of the node "sensor unit calls communication unit", including the influence of technical system components.

The CPTs of nodes that represent states of TSCs express the probability for a system component of either being in service (survival) at time t ($\Pr(S(t))$) or not being in service (failure) at time t ($\Pr(F(t))$). Failures occur following a Poisson process, i.e. occur randomly in time and independently of each other. The probability of failure at time t is given as

$$\Pr(F(t)) \approx \lambda \times \mathrm{E}[T_r], \qquad (1)$$

where λ is the failure rate and $\mathrm{E}[T_r]$ is the expected time it takes to detect and repair the failure. The approximation holds for small values of λ (Straub, 2012). The failure rate λ incorporates both internal technical failure rates λ_{IF} and failures caused by ex-

ternal influences λ_{EF} such as rock fall, lightning, animals, vandalism, humidity.

$$\lambda = \lambda_{IF} + \lambda_{EF,} \qquad (2)$$

Internal failure rates λ_{IF} for TSCs are directly derived from specified Mean Time To Failure (*MTTF*) data or, for repairable parts from Mean Time Between Failure (*MTBF*) data:

$$\lambda_{IF} = \frac{1}{MTTF} \quad \text{or} \quad \lambda_{IF} = \frac{1}{MTBF}, \qquad (3)$$

External failure rates λ_{EF} are derived from field data, from operating or maintenance records, incident data banks or expert opinions (Stewart & Melchers, 1997). Failures due to external causes are rare at Illgraben. In the period between 2006 and 2012, one out of two radar devices failed twice and one out of three solar panels once. One radar device was muddied during an event, measured unrealistic high values and continuously issued false alarms. Another time, the mounting of the radar device failed. The empirically observed external failure rates λ_{EF} for radar devices, based on two events for two sensors during a period of 7 years, is $\lambda_{EF} = 2 / (2 * 7 * 365) = 3.9 *10^{-4}$ per day. One solar panel was hit by a rock during a debris flow event and was almost completely destroyed. The resulting external failure rate λ_{EF} for the solar panel is $\lambda_{EF} = 1 / (3 * 7 * 365) = 1.3 *10^{-4}$ per day. By adding the external failure rates λ_{EF} to the specified internal failure rates λ_{IF}, the overall failure rates λ of the radar devices and solar panels increase significantly. For the other TSCs, the overall failure rate λ is equal to internal failure rate λ_{IF} and derived from *MTTF* or *MTBF* data, specified by the suppliers (Table 1).

Table 1. Input values to calculate the failure rate λ.

component	specified	λ_{IF}	λ_{EF}
geophone	100y MTTF*	2.7 *10^{-5}	no failure observed
radar	60y MTTF	4.5 *10^{-5}	3.9 *10^{-4}
data logger	180y MTBF	1.5 *10^{-5}	no failure observed
modem	125y MTBF	2.1 *10^{-5}	no failure observed
T-Box	45y MTBF	6.0 *10^{-5}	no failure observed
SMS Butler	50y MTBF*	5.4 *10^{-5}	no failure observed
audible alarm	20y MTTF*	1.3 *10^{-4}	no failure observed
red light	20y MTTF*	1.3 *10^{-4}	no failure observed
solar panel	50y MTTF*	5.4 *10^{-5}	1.3 *10^{-4}
battery	10y MTTF*	2.7 *10^{-4}	no failure observed

* assumptions.

It is estimated that the $\mathrm{E}[T_r]$ for all TSCs of the debris flow alarm system is about three days. Diagnosis tools are incorporated into the system to ensure that the detection and communication of failures takes place within one day. On the same day the spare parts are ordered. The delivery takes another day before the TSC is replaced on the third day.

For the mobile network and the public power supply, the probability of failure at time t, $\Pr(F(t))$, is specified in general. According to the local power

supplier, Illgraben was not supplied with power for 7.57 h between 2008 and 2011. It follows that the probability of failure for the power supply is $\Pr(F(t))$ = 0.315 d / 1825 d = 1.1 *10^{-4} per day. For the mobile network it is assumed (experience L. Meier, GEOPRAEVENT) that the probability of failure for the mobile connection is about 15 min per year. Thus the probability of failure per day is estimated at $\Pr(F(t))$ = 0.0104 d / 365 d = 2.8 *10^{-5} per day.

4.4 *The influence of the system threshold on the system reliability*

The overall system reliability depends not only on the reliability of the TSCs, but is also strongly influenced by the interpretation of the measurements and the choice of the system alarm threshold. These aspects are modeled directly in the BN. The final BN is designed in such a way that the system reliability in terms of the POD of an event versus the PFA can be calculated. Both probabilities are calculated automatically for different types of events (A / B / C / D / no event). The top node thus represents the natural hazard event type. If evidence is set on event A, B, C, or D the BN calculates the POD, while evidence set on "no event" calculates the PFA.

To analyze the influence of the threshold on the system reliability, sensor data, continuously measured in the period between 1st of May 2008 and 24th September 2012, are processed and interpreted within the BN. The analyzed period includes 6 A, 18 B, 2 C, 18 D events and 839 "no event" days. In the line below the top node "event", measurements of each of the five sensors are represented in independent nodes. Each CPT summarizes the frequency of occurrence of ground vibrations or flow depth differences measured by one specific sensor. These frequencies are further subdivided and the relative frequencies are stated conditional on the event type. To obtain the frequency of occurrence of ground vibrations, the maximum ground vibration value per day and geophone is identified. This maximum value

is a sufficient criterion to interpret the influence of the system threshold, because it indicates if an alarm is issued on a specific day. To obtain the difference in flow depth for a specific day, which is the alarm criterion for the radar, the respective maximum flow depth value of a given day is subtracted from the maximum value of the previous day. These maximum ground vibration values and flow depth differences for each day and sensor are then discretized in classes (Table 2).

Table 2. Ground vibration and flow depth classes.

ground vibration maximum		flow depth difference	
class	values (impulses / sec)	class	values (m)
GV1	= 0	FD1	≤ 0
GV2	> 0, ≤ 1	FD2	> 0, ≤ 0.1
GV3	> 1, ≤ 5	FD3	> 0.1, ≤ 0.3
GV4	> 5, ≤ 20	FD4	> 0.3, ≤ 0.7
GV5	> 20, ≤ 50	FD5	> 0.7, ≤ 1.5
GV6	> 50, ≤ 200	FD6	> 1.5, ≤ 3
GV7	> 200, ≤ 500	FD7	> 3
GV8	> 500		

To calculate the relative frequencies of occurrence for each sensor and event type, these classified maximum ground vibration and flow depth difference occurrence values are enumerated per event type and sensor. The resulting frequencies of occurrence for a given geophone and event type differ strongly (Figure 7). G1 records no ground vibrations on "no event" days that exceed GV4, whereas G2 and G3 monitor values in all classes. As a result, they are more likely to issue false alarms. On days with A events, G1 measures values in higher classes, between GV4 and GV6. The other geophones record values in lower classes for A events. G2 measures values between GV2 and GV5 and G3 between GV4 and GV8. Thereby, G1 and G3 follow logical patterns: The stronger the event type, the higher are the measured values. In contrast, G2 records values in lower classes even on days with strong events.

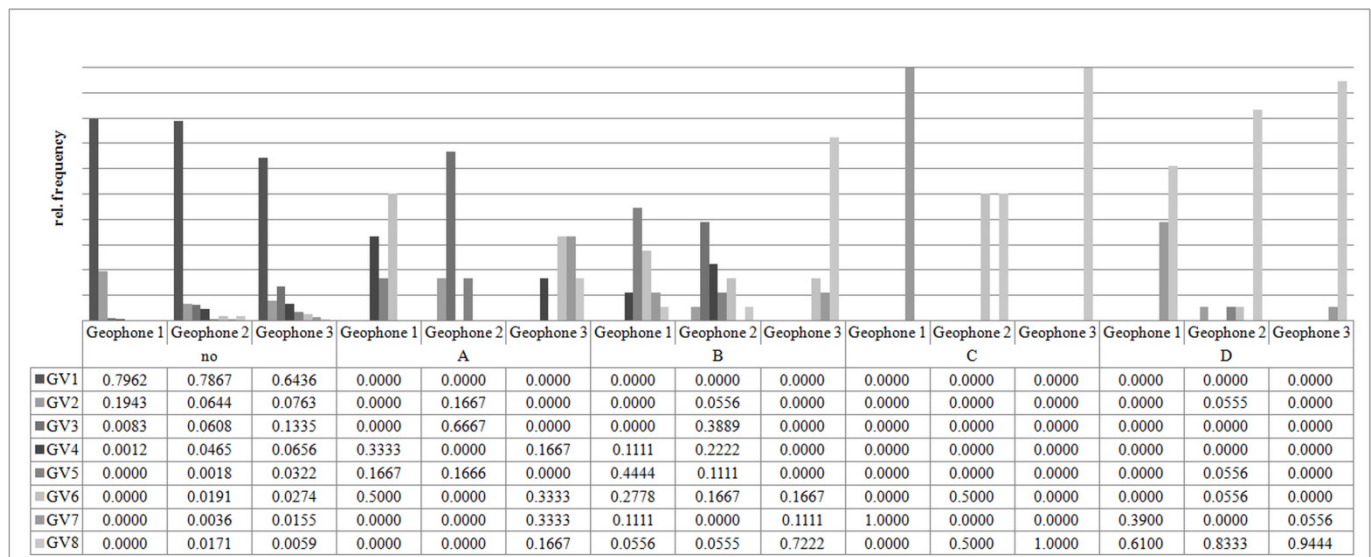|	Geophone 1	Geophone 2	Geophone 3	Geophone 1	Geophone 2	Geophone 3	Geophone 1	Geophone 2	Geophone 3	Geophone 1	Geophone 2	Geophone 3	Geophone 1	Geophone 2	Geophone 3
		no			A			B			C			D	
■GV1	0.7962	0.7867	0.6436	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
■GV2	0.1943	0.0644	0.0763	0.0000	0.1667	0.0000	0.0000	0.0556	0.0000	0.0000	0.0000	0.0000	0.0000	0.0555	0.0000
■GV3	0.0083	0.0608	0.1335	0.0000	0.6667	0.0000	0.0000	0.3889	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
■GV4	0.0012	0.0465	0.0656	0.3333	0.0000	0.1667	0.1111	0.2222	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
■GV5	0.0000	0.0018	0.0322	0.1667	0.1666	0.0000	0.4444	0.1111	0.0000	0.0000	0.0000	0.0000	0.0000	0.0556	0.0000
■GV6	0.0000	0.0191	0.0274	0.5000	0.0000	0.3333	0.2778	0.1667	0.1667	0.0000	0.5000	0.0000	0.0000	0.0556	0.0000
■GV7	0.0000	0.0036	0.0155	0.0000	0.0000	0.3333	0.1111	0.0000	0.1111	1.0000	0.0000	0.0000	0.3900	0.0000	0.0556
■GV8	0.0000	0.0171	0.0059	0.0000	0.0000	0.1667	0.0556	0.0555	0.7222	0.0000	0.5000	1.0000	0.6100	0.8333	0.9444

Figure 7. Measured ground vibration frequencies conditional on event type and geophone.
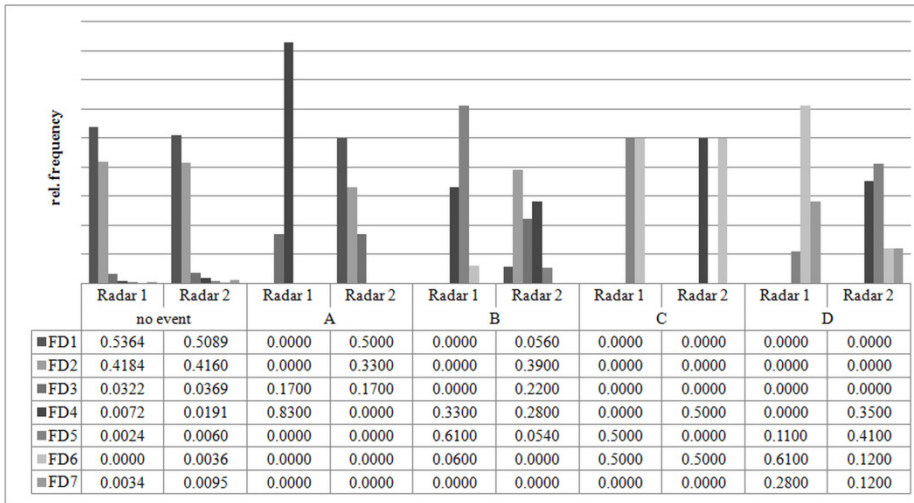
Figure 8. Measured flow depth difference frequencies conditional on event type and radar device.

	Radar 1	Radar 2	Radar 1	Radar 2	Radar 1	Radar 2	Radar 1	Radar 2	Radar 1	Radar 2
	no event		A		B		C		D	
■FD1	0.5364	0.5089	0.0000	0.5000	0.0000	0.0560	0.0000	0.0000	0.0000	0.0000
■FD2	0.4184	0.4160	0.0000	0.3300	0.0000	0.3900	0.0000	0.0000	0.0000	0.0000
■FD3	0.0322	0.0369	0.1700	0.1700	0.0000	0.2200	0.0000	0.0000	0.0000	0.0000
■FD4	0.0072	0.0191	0.8300	0.0000	0.3300	0.2800	0.0000	0.5000	0.0000	0.3500
■FD5	0.0024	0.0060	0.0000	0.0000	0.6100	0.0540	0.5000	0.0000	0.1100	0.4100
■FD6	0.0000	0.0036	0.0000	0.0000	0.0600	0.0000	0.5000	0.5000	0.6100	0.1200
■FD7	0.0034	0.0095	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2800	0.1200

Both radar devices measure flow depth differences in all defined classes on "no event" days and are thus prone to issue false alarms (Figure 8). As the event type gets stronger, R1 measures flow depth differences in higher classes, whereas data measured by R2 are difficult to interpret. On days with A and B events, R2 only measures small flow depth increases, similar to values measured on "no event" days. For stronger events R2 performs better and monitors significantly higher values.

4.4.1 POD and PFA of individual sensors

To determine the POD and the PFA for individual sensors, sensor units and the overall system, squared decision nodes are incorporated in the BN on two levels (Figure 9). Decision nodes, called "threshold", are incorporated to evaluate the influence of single system alarm thresholds on the sensors. The thresholds are chosen according to the defined classes (Table 1). The CPTs of their child nodes represent the terms and conditions set for an alarm release of a single sensor. An alarm is issued if the classified ground vibrations or flow frequencies of a certain event exceed the defined threshold.
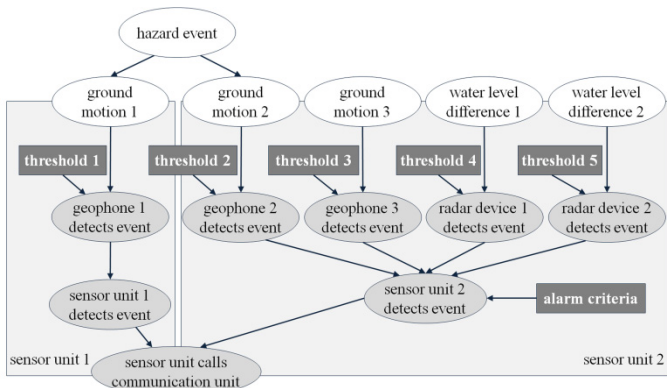

Figure 9. Decision nodes (squared) to determine the influence of the system thresholds.

The ROC curves for each individual sensor and event type vary (Figure 10). Independent of the event type,

the ROC curves of G1 are close to the optimum. Even for small events of type A, G1 is able to detect all events and reach a PFA of only 0.001 per day. Such an optimum threshold for the detection of A events does not exist for G2 and G3. If the threshold is defined in a way that all small events are to be detected with a high probability, an unacceptably high number of false alarms would result. For stronger events of the type C or D, a better ratio between POD and PFA can be reached, which is, however, still far from the optimum.
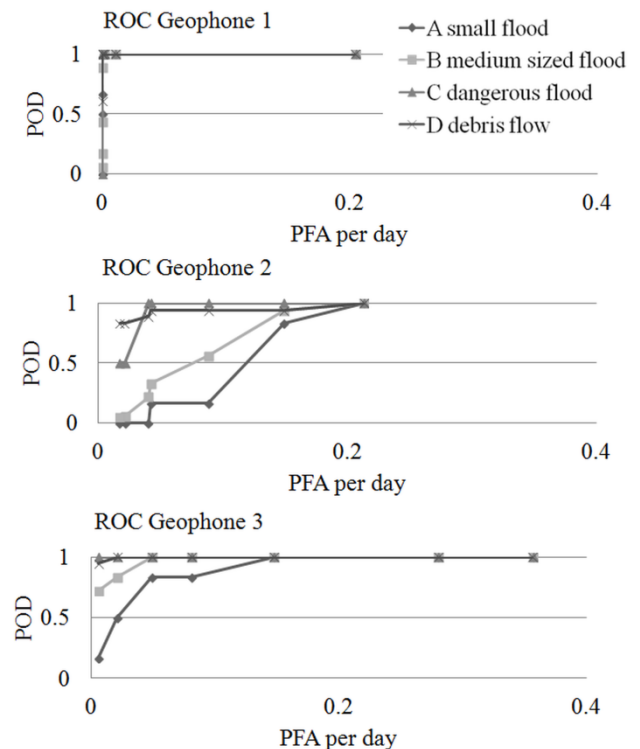

Figure 10. ROC curves for different event types and thresholds for geophone 1, 2 and 3, based on empirical estimates. Dots represent thresholds GV1 - GV8.

The performance of the two radar devices also varies according to the position and event (Figure 11). R1 performs much better, especially for the detection

of small events. For POD of values close to 1, the associated PFA is 0.045 per day. R2 reaches a maximum POD of 0.5 for small events, which implies a PFA of 0.5 per day. A better ratio between the POD and the PFA is only obtained by R2 for stronger events of type C and D.
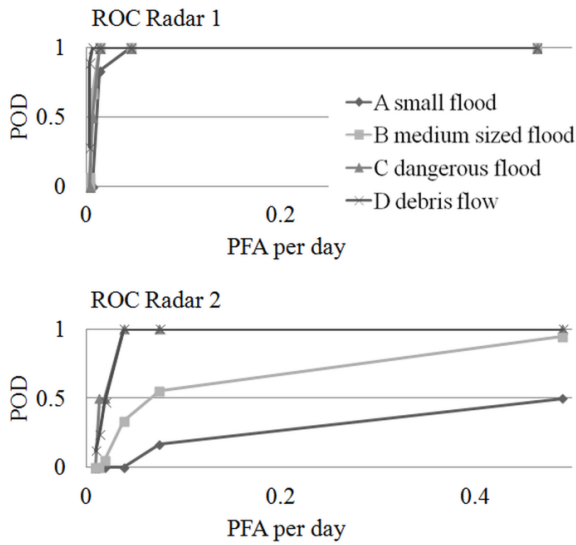

Figure 11. ROC curves for different event types and thresholds for radar 1 and 2, based on empirical estimates. Dots represent thresholds FD1 - FD7.

The thresholds implying the best ratio between POD and PFA for small events is identified from the ROC curves. For the only sensor in SU1 (G1), GV4 is optimal. For the other sensors two threshold combinations, T1 and T2, are considered. T1 includes thresholds that achieve a POD close to 1 (G2: GV2, G3: GV4, R1: FD3, R2: FD2) and T2 includes the thresholds values above T1 that imply lower POD values (G2: GV3, G3: GV5, R1: FD4, R2: FD3).

4.4.2 POD and PFA of the sensor unit 2
The good performance of SU1 is achieved by G1. The individual sensors in SU2 are prone to a high number of false alarms. To be able to combine alarm threshold criteria for the sensors in SU2 and reduce the PFA, the decision node "alarm criteria" is added. Through this decision node, four alarm criteria are defined. Criterion 1 (C1) specifies that each sensor can issue an alarm individually. Criterion 2 (C2) requires that the thresholds of at least one radar and one geophone are exceeded to issue an alarm. Criterion 3 (C3) specifies that at least three sensors must detect an event to issue an alarm and Criterion 4 (C4) requires that the thresholds of all four sensors are exceeded for an alarm. By means of the node "sensor unit 2 detects event", the POD or PFA for a chosen threshold combination (T1 / T2) and the alarm criterion (C1 - C4) in SU 2 can be calculated.

Independent of the threshold combination, an alarm combination C1 generates a maximum POD but unacceptably high values for PFA (Figure 12). The threshold combination T2 strongly reduces the

PFA. Threshold combinations play an important role in conjunction with C2 and C3. Only T1 in conjunction with C2 or C3 leads to a POD close to 1. POD values, generated through the threshold combination T2 and alarm criteria C2 or C3, are lower and C2 has a significant PFA. Alarm combination C4 implies a PFA equal to zero, but also a low POD.
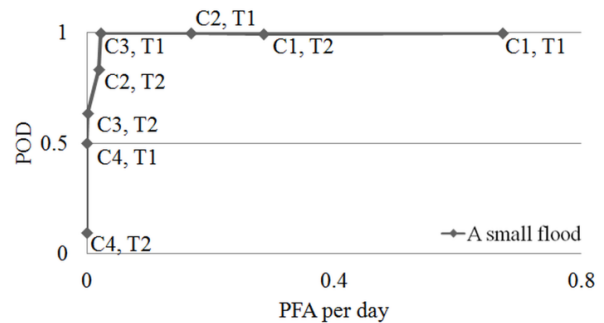

Figure 12. ROC curve for sensor unit 2 dependent on the threshold combinations (T1 or T2) and alarm criteria (C1 - C4).

4.4.3 POD and PFA of the overall system
The overall system reliability in terms of POD versus PFA can be calculated through the root node "alarm", by fixing the event type and computing the conditional probabilities of alarm. A high POD, between 0.998 and 0.999, is reached for all defined threshold combinations and alarm criteria of SU2 and the optimal threshold GV4 for G1 in SU1 (Figure 13). Despite the incorporation of the failure probabilities of all TSCs, the system achieves a maximum POD of 0.999. The PFA ranges between 0.6 and 0.001 per day. The best result leads to a POD of 0.999 and a PFA of 0.002 per day and is reached by alarm criterion C3 and, in contrast to section 4.4.2, by the threshold combination T2. The improved POD and decreased PFA values and the higher optimal threshold combination for SU2 are a consequence of the incorporation of an excellent performing G1 in SU1.
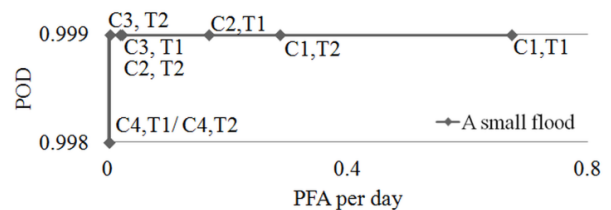

Figure 13. ROC curve for the alarm system dependent on the threshold combinations (T1 or T2) and alarm criteria (C1 - C4) and a threshold GV4 for the geophone in sensor unit 1.

5 DISCUSSION AND CONCLUSION

The case study shows that BNs are a suitable tool to analyze the reliability of warning and alarm systems for natural hazards. Uncertainties and system dependencies can be included. The conditional POD and PFA for the overall system, for individual sensors or sensor units can be calculated automatically for each event type.

The high overall system reliability observed for the Illgraben debris flow alarm system is reached by a redundant sensor configuration and the optimal performance of the single geophone in sensor unit 1. The analysis revealed that the choice of the thresholds and alarm criteria has the largest influence on the system reliability. Even identically constructed sensors have varying sensitivity to the hazard event. These differences are due to the positioning of the sensors, i.e. their proximity to the catchment. According to the analysis, sensors of sensor unit 2 perform less reliably. Even with combined alarm criteria, the PFA would be unacceptable high. However, our analysis overestimates the PFA of sensor unit 2. The reason is that the duration of signals is included as an additional alarm criterion. The threshold of a geophone is only exceeded if ground vibrations exceed the defined threshold for at least 5 seconds. Consequently, short excitations due to spontaneous events such as rock falls do not cause false alarms.

In this case study, failure rates of the technical system components have only a minor influence on the overall system reliability. The fact that a small number of external failures strongly increase the probability of failures for technical system components demands a detailed analysis in the future of external failure causes and associated probabilities. In addition, a cost-benefit analysis would give valuable information about the cost-effectiveness or redundant communication and alarm units.

The application of the BN and the resulting ROC curves showed that the selected methods are able to deal with criteria that influence the reliability of warning and alarm systems for natural hazards. A detailed sensitivity analysis is currently being conducted to identify the optimal threshold. The final reliability quantification method will include predefined acceptable risk levels and will explicitly quantify the consequences of false alarms and missed events as a basis for selecting the optimal warning strategy.

6 REFERENCES

Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R. & Salmerón, A. (2011) Bayesian networks in environmental modelling. *Environmental Modelling & Software,* vol. 26, no. 12, pp. 1376-1388.

Badoux, A., Graf, C., Rhyner, J., Kuntner, R. & Mcardell, B. W. (2009) A debris-flow alarm system for the Alpine Illgraben catchment: design and performance. *Natural Hazards,* vol. 49, no. 3, pp. 517-539.

Bensi, M., Kiureghian, A. D. & Straub, D. (2012) Efficient Bayesian network modeling of systems. *Reliability Engineering & System Safety,* vol. 112, pp. 200-213.

Blanchard, B. S. & Fabrycky, W. J. (2011) *Systems Engineering and Analysis,* New York, Pearson Education, Inc.

Bründl, M. & Heil, B. (2011) Reliability analysis of the Swiss avalanche warning system. IN Faber, M., Köhler, J. & Nishijima, K. (Eds.) *11TH International Conference on Applications of Statistics and Probability in Civil Engineering.* Zürich, CRC Press an imprint of the Taylor & Francis Group.

Bründl, M., Romang, H., Bischof, N. & Rheinberger, C. (2009) The risk concept and its application in natural hazard risk management in Switzerland. *Natural Hazards and Earth System Sciences,* vol. 9, pp. 801-813.

Dsl (2013) *GeNIe & SMILE,* Decision Systems Laboratory, University of Pittsburgh, viewed 25 March 2013 http://genie.sis.pitt.edu/index.php/about.

England, W. L. (1988) An exponential model used for optimal threshold selection on ROC curves. *Medical Decision Making,* vol. 8, no. 2, pp. 120-131.

Fawcett, T. (2006) An introduction to ROC analysis. *Pattern recognition letters,* vol. 27, no. 8, pp. 861-874.

Felgentreff, C. & Glade, T. (2008) *Naturrisiken und Sozialkatastrophen,* Spektrum Akademischer Verlag.

Friis-Hansen, A. (2001) *Bayesian Networks as a Decision Support Tool in Marine Applications,* Technical University of Denmark (DTU).

Graf, C., Badoux, A. & Mcardell, B. W. (2009) Alarmsystem Illgraben - Erfahrungen während der Pilotphase. *Wasser Energie Luft,* vol. 101, no. 2, pp. 101-107.

Intrieri, E., Gigli, G., Casagli, N. & Nadim, F. (2013) Brief communication "Landslide Early Warning System: toolbox and general concepts". *Nat. Hazards Earth Syst. Sci.,* vol. 13, no. 1, pp. 85-90.

Jensen, F. V. & Nielsen, T. D. (2007) *Bayesian networks and decision graphs,* New York, Springer Science + Business Media.

Medina-Cetina, Z. & Nadim, F. (2008) Stochastic design of an early warning system. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards,* vol. 2, no. 4, pp. 223-236.

Pagès, A. & Gondran, M. (1986) *System reliability: evaluation and prediction in engineering,* New York, Springer.

Sättele, M., Bründl, M. & Straub, D. (2012) A classification of warning system for natural hazards. IN Moormann, C., Huber, M. & Proske, D. (Eds.) *10th International Probabilistic Workshop.* Stuttgart, Institut für Geotechnik der Universität Stuttgart.

Schoefs, F., Clement, A. & Nouy, A. (2009) Assessment of ROC curves for inspection of random fields. *Structural Safety,* vol. 31, no. 5, pp. 409-419.

Stewart, M. G. & Melchers, R. E. (1997) *Probabilistic risk assessment of engineering systems,* London, Chapman & Hall.

Straub, D. (2005) Natural hazards risk assessment using Bayesian networks. IN Augusti, G., Schuëller, G. I. & Ciampoli, M. (Eds.) *9th International Conference on Structural Safety and Reliability ICOSSAR'05.* Rome, Millpress.

Straub, D. (2012) *Lecture Notes in Engineering Risk Analysis: Part E – Probabilistic models,* Technische Universität München.

Unep (2012) *Early Warning Systems: A State of the Art Analysis and Future Directions.* Nairobi, Division of Early Warning and Assessment (DEWA)