Quantifying the effectiveness of early warning systems for natural hazards

M. Sättele¹, M. Bründl¹, and D. Straub²

¹WSL Institute for Snow and Avalanche Research SLF, Davos Dorf, Switzerland
²Technische Universität München, Engineering Risk Analysis Group, Munich, Germany

Correspondence to: M. Sättele (saettele@slf.ch)

Received: 8 July 2015 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: 30 July 2015
Accepted: 17 December 2015 – Published: 19 January 2016

Abstract. Early warning systems (EWSs) are increasingly applied as preventive measures within an integrated risk management approach for natural hazards. At present, common standards and detailed guidelines for the evaluation of their effectiveness are lacking. To support decision-makers in the identification of optimal risk mitigation measures, a three-step framework approach for the evaluation of EWSs is presented. The effectiveness is calculated in function of the technical and the inherent reliability of the EWS. The framework is applicable to automated and non-automated EWSs and combinations thereof. To address the specifics and needs of a wide variety of EWS designs, a classification of EWSs is provided, which focuses on the degree of automations encountered in varying EWSs. The framework and its implementation are illustrated through a series of example applications of EWS in an alpine environment.

1 Introduction

A growing number of early warning systems (EWSs) is developed and operated for reducing the risks imposed by a wide range of natural hazard processes. They can mitigate the consequences of hazardous events if information is issued in a timely way. In recent years, EWS technologies have been improved significantly. In many fields, EWS are now cost-efficient alternatives to structural mitigation measures. They are applied for large-scale hazard processes, such as severe weather, floods, tsunamis, volcanic eruptions or wildfires, where they complement structural measures and support the preparation for, and response to, the hazard events (e.g., Sorensen, 2000; Zschau and Küppers, 2003; Grasso and Singh, 2009; Glade and Nadim, 2014). They are also popular as flexible and temporary mitigation measures on smaller scales. In mountain regions, they are successfully applied to mitigate risks from snow avalanches, debris flows, flash floods, rockfalls, and landslides (e.g., Bell et al., 2010; Thiebes, 2012; Michoud et al., 2013; Stähli et al., 2015).

Whether or not EWSs are effective and efficient risk mitigation measures can be evaluated case-specifically through cost–benefit analyses, in which the life-cycle costs and the efficiency are compared to those of alternative mitigation measures (Penning-Rowsell, 2005; SafeLand, 2012; Špačková and Straub, 2015). In cost–benefit analyses, the efficiency is defined as the risk reduction achieved with a mitigation measure and is expressed in monetary terms. To avoid expressing the risk in monetary terms, cost-effectiveness analyses can be conducted instead (Bründl et al., 2009). The effectiveness $E_w$ is quantifiable without expressing the risk in monetary terms. For EWSs, one can define it as a function of the overall risk without the EWS, $R$, and the risk with the EWS, $R^{(w)}$ (Sättele et al., 2015a):

$$E_w = 1 - \frac{R^{(w)}}{R}. \quad (1)$$

The risks with and without the EWS are evaluated by summing or integrating over all $n_{scen}$ possible scenarios $j$ and all $n_{obj}$ exposed objects $i$, which are persons or assets exposed to a hazardous scenario:

$$R = \sum_{j=1}^{n_{scen}} \sum_{i=1}^{n_{obj}} R_{ij}. \quad (2)$$

Both $R_{ij}$ and $R_{ij}^{(w)}$ can be calculated from the probability of occurrence of a hazard scenario, $p_j$, the probability of expo-
Figure 1. Following the principle of signal detection theory, a classifier (e.g. in form of a threshold) discriminates between correct and wrong outcomes of EWSs: the EWS correctly issues an alarm when an event occurs (hit) or no alarm when no event occurs (neutral), but can also wrongly issue false alarms or miss dangerous events.

\[
R_{ij} = p_j \times pe_{ij} \times v_{ij} \times A_i. \tag{3}
\]

When issuing timely information, EWSs can reduce the exposure probability of persons and mobile objects (Dai et al., 2002; SafeLand, 2012; Thiebes, 2012) or their vulnerability (Einstein and Sousa, 2006). Detailed guidelines on how this risk reduction can be evaluated have been published for structural mitigation measures (e.g. Romang, 2008) but, to the best of our knowledge, not for EWSs.

Even without detailed guidelines, the effectiveness of EWSs has been investigated previously. Thereby, it is common practice to consider both the probability that an EWS detects hazardous events, as well as the probability that the EWS leads to a false alarm. If the EWS detects a hazard event, timely warnings can initiate preventive actions, such as an evacuation of endangered persons to prevent damage. However, frequent false alarms can lead to excessive intervention costs or reduce compliance with future warnings (Pate-Cornell, 1986; Grasso et al., 2007; Schröter et al., 2008; Rogers and Tsirkunov, 2011; Ripberger et al., 2014). To account for the probability that events are correctly detected (hit) and the probability that false alarms are issued (Fig. 1), the effectiveness is typically evaluated based on concepts of signal detection theory, where a classifier (in the simplest case a predefined threshold) discriminates between alarm and no alarm (Swets, 1996).

An optimal EWS detects all hazardous events and never produces false alarms (Intrieri et al., 2013). In the operational application of EWS, false alarms cannot be avoided and an optimal trade-off between detected events and false alarms needs to be identified. To solve this optimization problem quantitatively, costs and utilities must be assigned to possible outcomes. Along these lines, Paté-Cornell (1986) suggests optimizing the effectiveness of fire warning systems operated in buildings in function of the probability that the event is detected (POD) and the probability that endangered persons comply with the warning (POC). The latter is modelled conditional on the probability of false alarms (PFA) by means of both descriptive (“How do people react in real situations?”) and normative (“How should people optimally react?”) approaches. In the normative model, the willingness of individuals to respond to an alarm is considered through a decision tree. Following that approach, decision trees have been used by others for the identification of decision rules that provide an optimal trade-off between POD and PFA (Einstein and Sousa, 2006; Rheinberger, 2013). In these two subsequent studies, the effect of false alarms on the compliance is not explicitly addressed, but the reliability is expressed in terms of POD and the PFA. This ability of the EWS to distinguish between hazard events and noise can be summarized graphically in receiver operator characteristic curves. This is the inherent reliability of an EWS and will be presented in Sect. 3.

As an alternative to decision trees, influence diagrams are applied to probabilistically model decision procedures associated with EWS (Einstein and Sousa, 2006; Martina et al., 2006). Influence diagrams are based on Bayesian networks (BNs), which are graphical models that consist of nodes representing random variables and arcs describing the stochastic dependencies among them (Jensen and Nielsen, 2007). They have been successfully applied in the field of environmental modelling and civil engineering due to their intuitive nature, their ability to deal with uncertainty and performing Bayesian analysis, and because of their strengths in representing dependence in large-scale systems (Straub, 2005; Straub and Der Kiureghian, 2010). Causal relations between components are defined through conditional probability tables, describing the probability distributions of the variables conditional on their parent nodes. Influence diagrams extend BNs for decision analysis by including decision nodes and utilities (Shachter, 1986).

In Sturny and Bründl (2014), a BN has been constructed to model the technical reliability of a glacier lake EWS. In their study, it was possible to model the entire technical system with a BN, which was not possible with a fault tree in a previous study on the reliability of the Swiss avalanche forecasting system (Bründl and Heil, 2011). The first BN, which models both the technical and the inherent reliability of a EWS, is described for a debris flow EWS in Sättele et al. (2015a). In a subsequent case study, the reliability of a partly automated rockslide warning system is assessed (Sättele et al., 2015b). The automated part is again modelled in a BN and human decision procedures of the non-automated part are assessed through a Monte Carlo analysis.

In the present contribution, a comprehensive framework approach for the evaluation of EWSs is presented, with three main objectives. The first objective, addressed in Sect. 2, is the development of a classification of EWSs, which serves as a basis for a structured evaluation of EWSs. The second objective is the development of evaluation methods for the tech-
Table 1. Characteristics associated with EWS classes.

<table>
<thead>
<tr>
<th>Alarm system</th>
<th>Warning system</th>
<th>Forecasting system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully automated</td>
<td>Partly automated</td>
<td>Lowest degree of automation</td>
</tr>
<tr>
<td>Detect ongoing process parameters</td>
<td>Monitor precursors</td>
<td>Monitor precursors</td>
</tr>
<tr>
<td>Short lead times</td>
<td>Extended lead times</td>
<td>Extended lead times</td>
</tr>
<tr>
<td>Thresholds serve as decision instance</td>
<td>First decision is based on threshold,</td>
<td>Experts conduct analysis in regular intervals and not based on thresholds</td>
</tr>
<tr>
<td>Automated intervention measures such as automated barriers on roads or interrupted power lines at railways</td>
<td>the final one is made by experts</td>
<td>Forecast the danger level for predefined warning regions to enable preventive actions and preparation</td>
</tr>
</tbody>
</table>

Figure 2. Classification of EWSs: each EWS class includes typical system components facilitating the monitoring and interpretation of data, and the dissemination of warnings. Automated system parts are highlighted in grey.

Generic classification of EWSs

EWSs can be defined as “sets of capacities needed to generate and disseminate timely and meaningful warning information to enable individuals, communities, and organizations threatened by a hazard to prepare and to act appropriately and in sufficient time to reduce the possibility of harm or loss” (UNISDR, 2007). EWSs currently operated in practice have widely varying designs, because they are preliminarily developed as prototypes to fit specific needs. They are ambiguously referred to as alarm, alert, detection, early warning, forecasting, monitoring, and warning systems. To facilitate a structured evaluation of EWSs, a recognized classification should be established.

A classification of landslide EWSs is proposed by Bell et al. (2010), in which monitoring systems, alarm, and expert systems are distinguished. We adapt this proposal by classifying EWS in function of their degree of automation into alarm, warning, and forecasting systems (Sättele et al., 2012). In Fig. 2, each system class is depicted with the three main units for monitoring, data interpretation, and dissemination. To indicate the degree of automation, components, which are operated automatically, are highlighted in grey.

In this classification, monitoring systems are not considered as a stand-alone class, because they do not actively issue warning information (Schmidt, 2002; Glantz, 2003). They are a central unit of every EWS, in which the environment is observed and relevant data are collected to increase the understanding of the processes. As proposed by Bell et al. (2010), alarm systems are understood as threshold-based fully automated EWSs. The term “expert system” is omitted because it is already used in the field of artificial intelligence to signify computer systems that imitate the decision abil-
Figure 3. Assignment of natural hazard processes to the proposed classification of EWSs: the system class depends on the availability and expressiveness of precursors and the available lead time.

2.1 Alarm system

Alarm systems are fully automated EWSs (Table 1; Fig. 2a). In the monitoring unit, sensors are installed to detect process parameters of already ongoing hazard events. They are primarily installed for processes triggered rather spontaneously, such as earthquakes, wildfires, tornados, small rockfalls and debris flows (Sättele, 2015). Thus, the remaining lead time is short and procedures include a minimal number of interfaces to ensure a reliable and fast information flow. Sensors are directly connected to a control tool, e.g. a data logger, in the interpretation unit. Here, data are analysed to issue and transfer automated warnings or to initiate mitigation actions when predefined thresholds are exceeded. Measured sensor data are transferred and stored in a central data management unit, which is commonly equipped with a diagnostics system. In the dissemination unit, automated intervention measures use optical signals or sirens to generate warnings. In some cases, power cut-offs are initiated to stop approaching trains. At the same time, risk managers and system operators receive information.

For example, a fully automated alarm system is operated to protect persons from debris flows within the Illgraben catchment in Switzerland (Badoux et al., 2009). One single geophone in the upper catchment and two geophones and two radar devices some hundred metres below should detect ongoing events in real time (Fig. 4). They measure the ground vibrations and the flow depth in the river bed. The upper geophone is controlled by one logger and another logger controls...
the remaining four sensors. An automated alarm is initiated if predefined thresholds are exceeded. The alarm information is transmitted via modem and communication devices to activate audible signals and red lights at three alarm stations. In parallel information is sent to system operators. The lead time of the alarm system is between 5 and 15 min.

2.2 Warning system

Warning systems are partly automated EWS (Table 1; Fig. 2b). In the monitoring unit, sensors or human observers monitor precursors of hazardous processes. Precursors are either events that trigger the hazard, such as intense rainfall, or relevant changes in the disposition that occur prior to the event. Therefore, warning systems are typically installed for natural hazard processes that evolve over time and provide precursors, such as tsunamis announced by earthquakes, volcanic eruption, or large-scale rockfalls (Sättele, 2015). Lead times are extended and enable a two-instance decision-making procedure in the interpretation unit. The first instance is automated: sensor data are transferred to a control tool that typically uses predefined thresholds to initiate automated warnings, similar to alarm systems. The warning is not directly issued to endangered persons but to experts, which constitutes the second decision instance. Experts analyse measured sensor data, and to predict the final event they often apply models or consult additional information sources, such as remote sensing data or reports from local observers. In the dissemination unit, organized intervention actions, such as evacuations and/or closures of roads and railway sections, are set up to mitigate the risk.

For example, in Preonzo, Switzerland, a warning system was installed to predict a mid-magnitude rockslide (Willenberg et al., 2009; Loew et al., 2012), which eventually occurred on 15 May 2012, with about 300 000 m$^3$ rock mass (Fig. 5). Five extensometers and a total station with 14 reflectors monitored increased displacement rates. In the automated part, warning information was sent when predefined thresholds were exceeded. In the non-automated part, displacement data were analysed by experts and the inverse velocity model was applied to predict the event timing, on the basis of which it was decided on further activities. Evacuations were ordered to protect the underlying factories and road. The available lead time is in the order of days.

2.3 Forecasting system

Forecasting systems have the lowest degree of automation (Table 1; Fig. 2c). In the monitoring unit, sensors or human observers monitor precursors to indicate the likelihood of dangerous events. They are chiefly operated to extend the short lead time achieved with alarm systems for spontaneous processes, such as severe weather, wildfires, or snow avalanches, but can also be found for processes that are more predictable such as rain-induced flood events (Sättele, 2015). In contrast to warning systems, the data interpretation is not initiated when predefined thresholds are exceeded, but it is conducted at regular intervals. Measured sensor data are
Figure 5. System sketch of the rockslide warning system in Preonzo including partly automated procedures in the monitoring, interpretation, and dissemination unit. (Figure based on pixmaps 2015 swisstopo (5704 000 000).)

Figure 6. System sketch of the national avalanche forecasting system in Switzerland including mainly non-automated procedures in the monitoring, interpretation, and dissemination unit. (Figure based on pixmaps 2015 swisstopo (5704 000 000).)
Reliability analysis
(automated part EWS)
Inherent reliability:
monitoring strategy, automated decision-instance, etc.
Technical reliability:
failure probabilities, configuration, etc.

Reliability analysis
(non-automated part EWS)
Inherent reliability:
human-decision making, model accuracy, monitoring strategy, etc.
Technical reliability:
failure probabilities, configuration, etc.

Effectiveness analysis
Risk reduction:
through timely detection (POD), etc.
Compliance reduction:
due to false alarm (PFA), insufficient lead time, etc.

Figure 7. The framework approach comprises three major parts that can be selected dependent on the EWS class to quantify the effectiveness as a function of the reliability.

Part I
Reliability analysis method
(automated part EWS)

Technical reliability:
- Draw system sketch
- Design BN
- Determine conditional probabilities
- Estimate component failure probabilities

Inherent reliability:
- Include sensor data and thresholds
- Quantify the reliability

Figure 8. Part I includes six steps to model the technical and inherent reliability of automated EWSs.

3 Framework for the evaluation of EWSs

Based on the classification, we suggest a framework for a structured evaluation of EWS effectiveness, consisting of three parts as illustrated in Fig. 7. For fully automated alarm systems, parts I and III are sufficient; for partly automated warning and forecasting systems, all three parts should be executed.

In parts I and II, reliability analyses are conducted, including the technical and the inherent reliability. The technical reliability analysis accounts for the availability of technical system components and their interdependencies in the system. The inherent reliability analysis differs for parts I and II. While the inherent reliability of automated EWSs (part I) depends on automated decision instances such as signal thresholds, non-automated EWSs (part II) rely primarily on human decision-making and the accuracy of models. In some...
cases, the model accuracy needs to be considered in part I as well, e.g. when earthquake alarm systems use models to detect events in real time. In both parts, the inherent reliability is expressed in terms of POD and PFA, as is the overall reliability.

In part III, the EWS effectiveness is quantified as a function of POD and PFA. The effectiveness is a direct function of POD, because timely detection leads to intervention measures that reduce consequences. A high number of false alarms may not only cause large costs for unnecessary interventions, but also decrease the probability that persons comply (POC). The POC is estimated from a basic compliance rate, combined with reduction factors to account for the effect of false alarms (PFA), insufficient lead time, and possibly other effects related to the communication and perception of the alarm/warning.

In the following, the three parts of the framework are summarized and individual steps are demonstrated with results of the two case studies Illgraben and Preonzo (Sättele et al., 2015a; Sättele et al., 2015b).

3.1 Part I: reliability analysis of automated EWS

In part I, the reliability achieved with fully automated alarm systems and the automated part of warning and forecasting systems is assessed in six steps (Fig. 8). Both the technical and inherent reliability are modelled together in a BN, which results in the POD and PFA of the automated system.

First, draw the system sketch. A system sketch is an essential basis to understand the EWS design and the dependencies among the components (see Figs. 4–6). It can be constructed according to the main units of an EWS and contains all main system components. The information flow is indicated by arcs and components are represented in form of squares or nodes. Redundant system parts are depicted redundantly in the sketch.

Second, design the BN. The basic BN can be derived from the system sketch. It consists of nodes and arcs, which can be structured according to the same three units (see Fig. 9). Oval nodes represent system components, and they are arranged according to the causal chain from the hazard event to the warning. This includes the main functionalities such as data measured, event indicated, warning issued, transmitted and released. Redundant system components and functionalities are also depicted redundantly in the BN. The arcs in the BN are directed to follow the information flow between functionalities and components. Decision nodes (squared nodes) are added in the BN to specify decision criteria on varying levels (see fifth step).

Third, determine conditional probabilities. Interrelations between the components and functionalities in the causal chain can be specified in conditional probability tables of oval nodes. In many instances, AND or OR relations are sufficient to describe the dependencies of individual components and functionalities, but any other type of logical or probabilistic relation can also be specified. AND relations represent serial connections, in which all components must work to ensure the underlying functionality; OR relations can be used to model redundant configurations.

Fourth, estimate component availabilities. The availability of individual components is specified in the conditional probability tables of oval nodes representing components. If the component can assume exactly two states (functioning or fail), the random variable is binary. If additional states are possible, these are specified in the conditional probability tables. Availability can often be derived from failure rates specified by the suppliers, to which one should add the rate of failures caused by external sources, such as extreme temperatures or disturbances due to human and animal activity.

Fifth, include sensor data and decision instances. Decision instances, such as warning thresholds, are added as squared decision nodes on various levels, either for single sensors or to specify warning criteria to combine information from several sensors. Probabilities of measured sensor data to exceed these criteria are included in the conditional probability tables of the nodes representing sensor signals. These probabilities are estimated conditional on the occurrence of an event. This fifth step is not necessary for forecasting systems which do not use automated decision instances.

Sixth, quantify the reliability. The last node of the causal chain (warning) is used to assess the overall reliability of the EWS. POD and PFA are obtained by changing the status of the top node (hazard event) and evaluating the BN. If the top node is set to “event”, the probability of the last node being in state “alarm” is equal to the overall system POD. Similarly, the PFA is obtained by setting the top node to “no event”. The same BN facilitates that the technical and the inherent reliability are assessed together or separately. To model the technical reliability alone, the status of the node “event indicated” is set to “yes”; to assess the inherent reliability, the status of all nodes representing technical system components is set to the state “functioning”.

Illustrative examples from the Illgraben and Preonzo case studies

The reliability of the fully automated Illgraben alarm system and the automated part of the Preonzo warning system is quantified following the six steps of part I (Fig. 8).

First, draw the system sketch. For the Illgraben and the Preonzo case study, system sketches are designed following the three main units for monitoring, data interpretation, and information dissemination, as shown in Figs. 4 and 5. The sketch only includes main components to keep the following steps manageable. For example, the data logger is considered together with the included software.

Second, design the BN. The BNs constructed for the Illgraben and Preonzo EWSs vary strongly. For the fully automated Illgraben debris flow alarm system, a comprehensive reliability analysis for the entire warning chain from the haz-
Table 2. The causal relations between functionalities and components are specified in the conditional probability tables of grey nodes. Here, two examples of deterministic nodes are shown. (a) OR logic of the redundant sensor units; (b) AND logic of sensors in monitoring unit 2.

<table>
<thead>
<tr>
<th>(a) Sensor unit 1 indicates event</th>
<th>yes</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor unit 2 indicates event</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Warning transmitted</td>
<td>yes</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>1</td>
</tr>
</tbody>
</table>

(b) Event indicated 1 (geophone 1)

<table>
<thead>
<tr>
<th>Event indicated 2 (geophone 2)</th>
<th>yes</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event indicated 3 (radar 1)</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Event indicated 4 (radar 2)</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Warning issued 2</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

Figure 9. The BN to model the overall reliability of the Illgraben alarm system is structured according to three main units. Grey nodes represent main functionalities in the causal chain; white nodes represent components; squared black nodes represent the decision instances on two levels, for details see Sättele et al. (2015a).
and event to warning is conducted as illustrated in Fig. 9. The inherent and the technical reliability are evaluated together and are expressed in terms of POD and the PFA. Grey nodes represent the causal chain and white nodes the components; thresholds are defined through the black decision nodes.

For Preonzo a simplified BN is constructed to model the ability of the system to provide timely warning information to decision-makers (Fig. 10). Here, the technical reliability alone is modelled, and sensor data and decision nodes are not included, so that the PFA cannot be computed here. This simplification is possible because warnings are sent directly to experts whose compliance should not be reduced by frequent warning information.

Third, determine conditional probabilities. In both BNs, the interrelations among system elements are specified either deterministically or stochastically in the conditional probability tables of grey nodes. In the causal chain of the Illgraben BN, warning information is transmitted if either sensor unit 1 or 2 issues an event (Table 2a), but the warning in sensor
unit 2 is only issued if at least one of the geophones and one radar device indicates an event (Table 2b).

Fourth, estimate component availabilities. In both case studies, availabilities of components are specified in the conditional probability tables of white nodes. All components can assume exactly two states; functioning and failed. For the Illgraben case study, availabilities $A$ of system components are calculated following Eq. (4) and are in the order of $0.9995$ for most components (Sättele et al., 2015a).

$$A \approx 1 - (\lambda_{IF} + \lambda_{EF}) \times E[T_i]$$

(4)

$\lambda_{IF}$ are internal failure rates and $\lambda_{EF}$ are external failure rates; $E[T_i]$ is the expected time it takes to detect and repair a failure. Internal failures rates $\lambda_{IF}$ are derived from the specified mean time to failure (MTTF) and the mean time between failure (MTBF) values, and external failure rates $\lambda_{EF}$ are estimated by experts.

Fifth, include sensor data and decision instances. In the Illgraben case study, past event data from 44 events are used to determine probabilities of thresholds being exceeded on both event and non-event days (see Table 1 in Sättele et al., 2015a). The BN constructed for the warning system in Preonzo is developed to facilitate the assessment of the technical reliability alone and does not include thresholds or measured sensor signals (for details, see second step).

Sixth, quantify the reliability. In the Illgraben case study, the inherent reliability for varying thresholds is modelled for each sensor separately (see Fig. 11). Besides the threshold, the positioning of the sensors has a major influence on the EWS reliability, whereas technical failures of individual components have a comparatively low impact due to high redundancies (Sättele et al., 2015a).

For Preonzo we find that the technical reliability, i.e. the POD of the automated part, is high (0.988) due to multiple redundancies in the sensor unit and a diagnostic system that immediately detects and reports component failures to minimize downtimes of the system. The inherent reliability is close to 1, but is not assessed quantitatively with the BN. This is not necessary because the warning thresholds were set low to ensure that the EWS sends timely information to the expert team responsible for the final decision on an evacuation. The system is furthermore designed as fail-safe, i.e. in case of a technical failure, the experts are alerted.

### 3.2 Reliability analysis II: non-automated EWS

In part II, reliability analyses of non-automated parts of warning and forecasting systems are conducted. Here, the ability of the decision-makers to correctly predict or forecast events is evaluated. This ability depends on (potentially complex) human- and model-based decision procedures, which are difficult to quantify in practical applications. If the reliability cannot be expressed quantitatively in terms of POD and PFA, a qualitative or semi-quantitative analysis should be conducted instead. This evaluation should address both the technical and the inherent reliability and can be conducted in five steps (Fig. 12).

First, determine the minimal required lead time. Lead times associated with the non-automated part of warning and forecasting systems are typically larger than those of alarm systems, often in the range of 1 to several days (see Sect. 2.2). During this time period, additional data and information are collected and predictions become increasingly accurate (see e.g. Grasso et al., 2007; Schröter et al., 2008). The reliability analysis in part II is therefore conducted as a function of the lead time. The reliability can either be evaluated for a fixed lead time or for a set of lead times. For a given lead time, one should consider the reliability associated with that lead time, as well as the related intervention costs, e.g. those caused by an early evacuation.

Second, estimate failure probabilities of remote components. Non-automated EWSs measure precursors and thus provide extended lead times. Nevertheless, their reliability increases with shorter lead times. For some EWSs, destructive pre-events can lead to an increased failure probability of system components, e.g. sensors, as the event approaches. A typical example is provided by the Preonzo case study, summarized in Sect. 3.2.1. The technical failure probability associated with the minimum required lead time is the input for determining the remaining number of sensors, which will in turn affect the forecast accuracy that is evaluated in the next step.

Third, estimate model accuracy. Experts often apply models to predict the event magnitude, time, and spatial dimensions. Flood forecasts are for example based on coupled hydrometeorological models, which become probabilistic when Hydrological Ensemble Prediction Systems are used (Wet-
The accuracy of models depends on their capabilities, their case-specific applicability, and on the quality of the available input data. The quality of the data is determined by the number, the type, and the positioning of sensors. The model accuracy is evaluated for the selected minimal lead time and expressed qualitatively or semi-quantitatively (see fifth step). The estimated model accuracy directly influences the ability of decision-makers to set up intervention measures correctly. If no models are applied, this step can be skipped.

Fourth, evaluate human decision-makers. In the non-automated part of EWSs, the final decision is made by humans. The decision procedures involved are typically complex and can only be assessed quantitatively in some cases (see Sect. 3.2.1). In most cases, a qualitative or semi-quantitative analysis is more suitable, in which possible outcomes, the degree of risk aversion, and the expertise of individuals play a role.
individuals and effects associated with group dynamics are addressed. Decision-makers are evaluated according to their ability to correctly detect dangerous events (POD) and avoid false alarms (PFA). Both terms can be rated in predefined evaluation scales e.g. as low, medium, or high.

Fifth, evaluate the reliability. The reliability achieved in the non-automated part of the EWS is evaluated as a function of the lead time. It depends on the procedures to initiate and carry out intervention measures following a warning, such as evacuation. The decision on a warning is influenced by the accuracy of the applied forecasting models and the quality of available information from different sources, such as measured sensor data, data from other sources, and reports from human observers. The quality of the input information directly influences the forecast ability of models and the success of human decision-making. Whether damage is successfully prevented depends also on the quality and the feasibility of predefined intervention plans. In a comprehensive reliability analysis, all these factors and their dependencies are considered. In most cases, this analysis will be qualitative. However, the final reliability should be expressed (semi-)quantitatively in terms of POD and PFA. To this end, values for POD and PFA may be assigned to qualitative rating scales (e.g. low (POD = 0.90 and PFA = 0.1), medium (medium POD = 0.95 and PFA = 0.05), and high (POD = 0.99 and PFA = 0.01).

3.2.1 Illustrative example from the Preonzo case study

In a detailed case study, the reliability of the non-automated part of the Preonzo warning system is assessed. To enable a quantitative reliability evaluation, a post-event analysis of a large event (about 300,000 m$^3$) that occurred on 15 May 2012 is conducted, following the five steps of part II.

First, determine the minimal required lead time. If decision-makers release the information 1 day in advance, the evacuation can be carried out successfully and sufficient time for intervention teams to set up protective measures is available. The quality of the prediction is also maximum for short lead times, and the intervention costs, which occur due to business interruptions in the underlying factory buildings, can be kept relatively low. Hence, 1 day is selected as the lead time.

Second, estimate failure probabilities of remote components. Sensors fail before the event in May 2012, and shortly after the instable mass collapses, the majority of sensors are destroyed. To account for the increasing failure rate, a function is fitted to the number of observed failures (Fig. 13). The estimated failure probability of sensors at the minimal required lead time ($t = 1$ day) necessary to set up an evacuation successfully is 0.4.

Third, estimate model accuracy. To predict the event time, the inverse velocity model is applied on sensor data measured in Preonzo before 15 May. In Fig. 14, the predicted event dates modelled between 1 April and 14 May by sensors installed close to the release area are summarized. As the event approaches, the prediction made by individual sensors becomes more uniform. One day before the event occurred, at the minimal lead time, 10 out of 12 available sensors predict the event to occur on the next day. However, on 6 May, most sensors predict the event for the next day and an unnecessary evacuation is set up on May 7 and annulled a day later when accelerations slow down again.

Fourth, quantify human decision-makers. In Preonzo, the final decision on setting up intervention measures is made by an expert team. As a first attempt to quantify the decision-making procedure, the experts are characterized by simple decision rules. According to these rules, an evacuation is set up if less than a certain number of initial sensors remain intact (technical criterion) or if a certain percentage of initial sensors predict the event for the following day (inherent criterion), as summarized in Table 3. The number of initial sensors is varied in the Preonzo study from 5 to 50.

Fifth, quantify the reliability. The overall reliability achieved in the non-automated part of the Preonzo warning system is assessed probabilistically through a Monte Carlo simulation. The model accuracy and the sensor failures are randomized, to quantify the probability that evacuation measures are set up on the day of the event (POD) (Fig. 15a). In addition, the costs for intervention are calculated, which decrease with an increasing number of sensors, and which are smaller for the risk-tolerant decision-maker (Fig. 15b).

Analyses are conducted for a varying number of initial sensors and two risk types (see Table 3) and confirm that the risk tolerance of human decision-makers have a significant influence on the reliability of non-automated parts of EWS. Figure 15a shows that even with a high number of sensors, the probability of the risk-tolerant decision-maker detecting the event never exceeds 0.85.

3.3 Part III: effectiveness analysis

The effectiveness of an EWS, $E_w$, is defined as the relative risk reduction achieved with the EWS and can be quantified following Eq. (1) as a function of the risk without the EWS $R$ and the risk with the EWS $R^{(w)}$. EWSs reduce the risk when timely information leads to intervention measures that decrease either the exposure probability $pe_{ij}$ or in some cases the vulnerability in Eq. (3). By combining Eqs. (1)–(3), the effectiveness of an EWS can be calculated as

$$E_w = 1 - \frac{\sum_{j=1}^{n_{scen}} \sum_{i=1}^{n_{obj}} p_j \times pe_{ij}^{(w)} \times v_{ij}^{(w)} \times A_i}{\sum_{j=1}^{n_{scen}} \sum_{i=1}^{n_{obj}} p_j \times pe_{ij} \times v_{ij} \times A_i}.$$

To determine $pe_{ij}^{(w)}$ and $v_{ij}^{(w)}$, the POD and PFA, estimated in the reliability analyses of part I and II, are used.

The exposure probability $pe_{ij}^{(w)}$ is reduced when persons are successfully evacuated or when intervention measures...
Table 3. To quantify the human decision-maker, two risk types are specified with different evacuation criteria (Sättele et al., 2015b).

<table>
<thead>
<tr>
<th>Risk type</th>
<th>Technical evacuation criterion, evacuate when</th>
<th>Inherent evacuation criterion, evacuate when</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less risk-tolerant</td>
<td>fewer than 6 sensors are functioning</td>
<td>20 % of sensors forecast the event for the next day</td>
</tr>
<tr>
<td>More risk-tolerant</td>
<td>fewer than 3 sensors are functioning</td>
<td>50 % of the sensors forecast the event for the next day</td>
</tr>
</tbody>
</table>

Figure 15. The reliability (POD) and costs for intervention are modelled for two decision-makers and a varying number of initial sensors: (a) the less risk-tolerant decision-maker reaches high values of POD independent of the number of sensors; the risk-tolerant decision-maker only reaches a POD up to 0.85; (b) the more risk-tolerant decision-maker creates lower expected costs, which reach a minimum of CHF 215 000 with around 20 sensors or more; for details see Sättele et al. (2015b).

Figure 16. Compliance frequency as a function of the false alarm ratio (Sättele et al., 2015a).

Figure 17. The effectiveness of the Illgraben alarm system could be quantified as a function of POD and PFA; i.e. the reliability (Sättele et al., 2015a).

The vulnerability $v^{(w)}_{ij}$ is reduced if the EWS sends timely information that leads to temporary measures, which decrease the susceptibility of objects to damage. If storm events are announced in a timely way, movable objects can be fixed; if flood warnings are issued, protective temporary measures...
such as sandbags or wooden barriers can be installed. Modern earthquake alarm systems can slow down trains or shut down critical processes in factories when strong shaking is detected in time.

The reduction of the exposure probability and the vulnerability is equal to the probability that the event is detected and intervention measures are initiated (POD) and that endangered persons comply with the warning (POC). The latter is not relevant for fully automated intervention measures such as power cut-offs. If EWSs issue warnings to persons, a high POC is crucial. It can be quantified as a function of the general compliance rate $\text{POC}_0$ and reduction factors $\text{RF}$, e.g. due to false alarms $\text{RF}(\text{PFA})$ or insufficient lead time $\text{RF}(\text{ILT})$:

$$\text{POC} = \text{POC}_0 \times \text{RF}(\text{PFA}) \times \text{RF}(\text{ILT}).$$

(6)

The basic compliance rate and the reduction factors must be determined case-specifically. The basic compliance rate depends on the type of intervention measures and human decision-making. If, for example, barriers are closed on a road, car drivers have to comply, while red lights can be ignored. Moreover, it can be assumed that regular training and education leading to a higher awareness of potential consequences can improve the basic compliance rate.

The reduction factor due to false alarms $\text{RF}(\text{PFA})$ accounts for the cry-wolf effect, namely that people have an increased tendency to ignore warnings after experiencing (multiple) false alarms. This effect depends, among other factors, on past experiences, expected consequences, and the degree of risk aversion of the recipients.

The reduction factor due to insufficient lead time $\text{RF}(\text{ILT})$ expresses the ability to comply. In certain cases, EWSs have to be constructed in a way that the available lead time may not be sufficient and not everybody willing to comply can successfully evacuate. In the case of earthquake alarm systems, lead times are in the range of just a few seconds; or for avalanche alarm systems constructed above railways, the lead time is limited by the distance from the railway to the release point.

### 3.3.1 Illustrative example from the Illgraben case study

In the Illgraben case study, the effectiveness $E_w$ is calculated as a function of POD and PFA. The alarm system reduces the exposure probability of persons in the Illgraben catchment. Therefore, the effectiveness is equal to the reduced exposure probability with the EWS. To simplify the analysis, different debris flow types are not distinguished, and only one scenario $j$ is considered. The exposure probability is the same for all persons $i$, $pe_{ij} = pe_j$, and it follows

$$E_w = 1 - \frac{p_j \times pe_{j}^{(w)} \times \sum_{i=1}^{n_{\text{pers}}} v_{ij} \times A_i}{p_j \times pe_j \times \sum_{i=1}^{n_{\text{pers}}} v_{ij} \times A_i} \equiv 1 - \frac{pe_{j}^{(w)}}{pe_j}. \quad (7)$$

The reduced exposure probability is evaluated as a function of the POD and the POC:

$$pe_{j}^{(w)} = pe_j (1 - \text{POD} \times \text{POC}). \quad (8)$$

Inserting in Eq. (7), the effectiveness becomes

$$E_w = \text{POD} \times \text{POC}. \quad (9)$$

POD values result from the reliability analysis and POC is calculated as a function of PFA. To this end, we adapt the basic compliance rate $\text{POC}_0 = 0.95$ from published traffic analyses (Rosenbloom, 2009; Johnson et al., 2011) and the RF(PFA) is adapted from a existing case study in which the compliance frequency of students as a function of false alarms is assessed (Bliss et al., 1995). As illustrated in Fig. 16, the compliance frequency strongly decreases with an increasing ratio of false alarms.

In the Illgraben case study we extend the BN to a decision graph and identify the threshold combination that leads to a maximal effectiveness following Eq. (9). In Fig. 17, the resulting effectiveness is shown as a function of POD and PFA, together with the POD and PFA values associated with the best system configurations. For this highly reliable EWS, the effectiveness decreases faster with increasing PFA than with increasing POD.

### 4 Discussion

The proposed classification of EWSs distinguishes alarm, warning, and forecasting systems according to their degree of automation, their lead time, and the expressiveness of the available precursors (Figs. 2 and 3). The selection of an EWS class depends strongly on the underlying natural hazard process. Different process types allow for different monitoring strategies, which are associated with different lead times and degrees of automation. Earthquakes, for example, occur without clear precursors and damage can only be reduced by fully automated alarm systems with very short lead times. In contrast, large river floods provide clear precursors and damage can be reduced when warnings or forecasts are made early enough to set up temporary intervention measures.

A differentiation of EWSs according to their degree of automation has proven to be a valuable basis for evaluating EWSs. The system requirements differ strongly between automated and non-automated EWSs and these should be addressed separately. Typical procedures conducted within automated EWS parts are less complex than human- and model-based decision procedures that are part of non-automated EWSs. Part I of the proposed framework consists of a six-step method for a quantitative reliability assessment of automated EWS; and part II contains five steps for a qualitative or semi-quantitative evaluation of non-automated parts.

Through the two case studies, we demonstrate that this framework approach is applicable to assess alarm and warn-
ing systems installed for gravitational processes in mountain regions. With the Preonzo case study, we moreover show that under some conditions the reliability of non-automated EWSs can be quantified as well. Here, a post-event analysis is conducted, in which human decision-makers are specified through simple decision rules. When specifying less risk-tolerant decision rules (Table 3), the analysis leads to similar recommendations than the ones that were actually made by the experts. However, to refine the framework approach for the application to EWSs operated for earthquakes, floods, meteorological hazards, tsunamis, volcanic eruptions, and wildfires, the following steps of the procedure should be further enhanced.

In part I, the technical and the inherent reliability of automated EWSs are quantified in a BN. For the construction of the BN, a system sketch forms the basis for understanding key system components and their interrelations. To keep the complexity of the BN and the proceeding steps low, only essential components should be considered. In step four, availability of individual system components are estimated. Internal failure rates can be derived from specifications of manufacturers, but external failure sources such as extreme temperatures and lightning, which are more difficult to estimate, must be considered as well. However, for many EWSs such as the Iligraben case study, the influence of technical reliability is low compared to the inherent reliability, i.e. the ability to interpret data correctly. The assessment of the inherent reliability is challenging in the design phase of EWSs or for EWSs installed for rare events such as large-magnitude rockfalls. In these cases, sensor data are not yet available to estimate probability distributions of EWS signals. Other EWSs, such as earthquake alarm systems, use real-time models to estimate the magnitude on a spatial dimension whenever unexpected ground shaking is detected. Here, measured signals are vector values and vary in space and time; they need to be further processed in models before a classifier can be applied to distinguish critical events from non-occurrences. In these instances, BN must be enhanced, e.g. to model the reliability dependent on the lead time.

In part II, a qualitative or semi-quantitative evaluation is proposed, to assess time-dependent human- and model-based decision procedures. Although a concrete evaluation method, such as the BN of part I, is not provided, the overall procedure for the evaluation of non-automated EWSs is presented. The reliability is estimated as a function of the lead time. In step two, the increase in sensor failure probability before the event must be addressed, as demonstrated in the Preonzo case study. Another example is provided by the 2011 Tohoku earthquake in Japan 2011, where a majority of the offshore sensors failed before the tsunami hit the mainland (Wei et al., 2013). It may be possible that no sensor data are available for an event prediction in the critical phase. The accuracy of predictive models (step three) depends on the capacity of the model, its applicability, and the availability of sensor data. For natural hazard EWSs, it is common practice to express the accuracy of models in terms of POD and PFA (see Simmons and Sutter, 2009). As we demonstrate, the framework enables the possibility of technical system component failures to be included into POD and PFA, to obtain a single measure of EWS reliability. In some cases, e.g. for flood models, the ability to spatially and temporarily predict the event should be addressed in the reliability analysis (Wheater et al., 2005). In these cases, the reliability is ideally described by the prediction errors of the timely forecasted discharge and not (only) in terms of POD and PFA. In non-automated EWS, the final decision is made by humans, often together with models applied on available sensor data. In most cases, human decisions are not rule-driven and cannot be quantified easily, but depend on factors such as experience, risk tolerance, and the environment in which the decision is made. To account for those factors, a qualitative evaluation is proposed, in which the performance of human decision-makers is rated on predefined scales (e.g. low, medium, high) as is common for the evaluation of structural mitigation measures (Marrgheth and Romang, 2010). The final reliability should then be evaluated in a semi-quantitative procedure where values for POD and PFA are assigned to different rating scales, e.g. high POD (0.95–1.0), limited POD (0.8–0.95), and low POD (0–0.8).

In part III, the effectiveness is quantified as a function of POD and PFA. The reduction of the exposure probability and vulnerability is a direct function of POD. In some instances, the EWS effectiveness is directly proportional to POD, as demonstrated in the Iligraben case study. The PFA determines the probability that persons comply with the warning (POC). It is also used to estimate the costs caused by unnecessary evacuations. The costs and the effectiveness are main criteria for the identification of optimal risk mitigation measures for natural hazards.

The overall user-friendliness of the novel framework can be improved if a convenient software tool is provided. Such a software tool can be developed following the three steps defined by the framework approach. The reliability evaluation for automated system parts can be done by running a BN in the background. The user interface should be designed to be user-friendly, including simple input fields in which e.g. system components, their technical failure probabilities, and their dependencies can be specified in order to optimize a system. Finally, it could be embedded in a software environment in which risk reduction of an EWS can be compared to alternative measures to support decision-makers in the identification of optimal mitigation measures.

5 Conclusion

With the proposed framework approach, the effectiveness of EWSs is evaluated as a function of the reliability through three main parts. To enable a structured evaluation of EWS, a generic classification is provided, differentiating EWSs into
alarm, warning, and forecasting systems according to their
degree of automation, lead time, and the availability of clear
precursors. In function of the EWS class, different parts of
the framework are selected. Each part is structured along
predefined steps, which are illustrated here by the results of
two case studies. The reliability assessment of the automated
part of EWS is performed quantitatively through a Bayesian
network. To evaluate non-automated EWS parts, which in-
volve the decision-making of experts, a qualitative or semi-
quantitative approach is generally preferable. However, as
exemplified in the Preonzo case study, a quantitative assess-
ment is possible and provides insights.

The framework should be tested and further developed
through additional case studies. Findings of these studies can
be implemented in the existing approach, which is flexible
enough to cover various needs.

Acknowledgements. This framework is a main result of the project
ReWarn, which was financially supported by the Swiss Federal
Office for Civil Protection (FOCP). We want to thank Christoph
Graf (Swiss Federal Institute for Forest, Snow, and Landscape
Research WSL), Giorgio Valenti, and Marco Franzi (Canton
Ticino) as well as Lorenz Meier (Geopraevent AG) for their support
in the case studies. We also thank Unni Eidsvig (NGI) and an
anonymous reviewer for their valuable comments, which helped to
improve an earlier version of the paper.

Edited by: N. Kerle
Reviewed by: U. Eidsvig and one anonymous referee

References

Badoux, A., Graf, C., Rhyner, J., Kuntner, R., and McArdell, B. W.: A
Bell, R., Mayer, J., Pohl, J., Greiving, S., and T. G.: Integrative Frühwarnsysteme für gravitative Massenbewegungen (ILEWS):
Monitoring, Modellierung, Implementierung, KlarText Verlag, 
Essen, 270 pp., 2010.
Bliss, J. P., Gilson, R. D., and Deaton, J. E.: Human probability
matching behaviour in response to alarms of varying reliability,
Bründl, M. and Heil, B.: Reliability analysis of the Swiss avalanche
warning system, in: 11TH International Conference on Applications of Statistics and Probability in Civil Engineering, edited by:
Dai, F., Lee, C., and Ngai, Y. Y.: Landslide risk assessment and
Einstein, H. H. and Sousa, R.: Warning systems for natural threats,

Fuchs, S.: Cost-Benefit Analysis of Natural Hazard Mitigation, in:
Glade, T., and Nadim, F.: Early warning systems for natural hazards and
risks, Nat Hazards, 70, 1669–1671, 2014.
Glanz, M. H.: Usable science 8: early warning systems: do’s and
don’t’s, Report of workshop, Shanghai, China, 76 pp., 2003.
Analysis and Future Directions, Division of Early Warning and
Assessment (DEWA), United Nations Environment Programme
(UNEP), Nairobi, 2009.
Grasso, V. F., Beck, J. L., and Manfredi, G.: Automated decision
Hazards, IDNDR Early Warning Program, 35 pp., 1997.
Hess, J. and Schmidt, F.: Towards optimised early warning develop-
Intieri, E., Gigli, G., Casagli, N., and Nadim, F.: Brief com-
munication “Landslide Early Warning System: toolbox and
general concepts”, Nat. Hazards Earth Syst. Sci., 13, 85–90,
Jackson, J.: Introduction to expert systems, Addison-Wesley Long-
Jensen, F. V. and Nielsen, T. D.: Bayesian networks and decision
graphs, 2 ed., Information Science and Statistics, edited by: Jor-
dan, M., Kleinberg, J., and Schölkopf, B., Springer Science +
Johnson, M., Newstead, S., Charlton, J., and Oxley, J.: Riding
through red lights: The rate, characteristics and risk factors of
non-compliant urban commuter cyclists, Accident Analysis Pre-
vention, 43, 323–328, 2011.
Loew, S., Gischig, V., Moore, J., and Keller-Signier, A.: Monitoring
of potentially catastrophic rockslides, Proc. of 11th Int. and 2nd
North Am. Symp. on Landslides and Engineered Slopes, Banff,
Canada, 2012.
Margreth, S. and Romang, H.: Effectiveness of mitigation measures
against natural hazards, Cold Regions Science and Technology, 64, 199–207, 2010.
Martina, M. L. V., Todini, E., and Libralon, A.: A Bayesian deci-

dion approach to rainfall thresholds based flood warning, Hydrol.
Earth Syst. Sci., 10, 413–426, doi:10.5194/hess-10-413-2006,
2006.
Meister, R.: Country-wide avalanche warning in Switzerland, Pro-
Michoud, C., Bazin, S., Bikra, L. H., Derron, M.-H., and Jaboyed-
off, M.: Experiences from site-specific landslide early warn-
ning systems, Nat. Hazards Earth Syst. Sci., 13, 2659–2673,
Pate-Cornéll, M. E.: Warning Systems in Risk Management, Risk
Analysis and Future Directions, Division of Early Warning and
Assessment (DEWA), United Nations Environment Programme
(UNEP), Nairobi, 2009.
Rheinberger, C. M.: Learning from the past: statistical perfor-


