



# Proportional loss functions for debris flow events

C. M. Rheinberger<sup>1</sup>, H. E. Romang<sup>2</sup>, and M. Bründl<sup>3</sup>

<sup>1</sup>LERNA-INRA, Toulouse School of Economics, Toulouse, France

<sup>2</sup>Swiss Federal Office of Meteorology and Climatology, Zurich, Switzerland

<sup>3</sup>WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

Correspondence to: M. Bründl (bruendl@slf.ch)

Received: 18 December 2012 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: –  
Revised: 25 June 2013 – Accepted: 8 July 2013 – Published: 30 August 2013

**Abstract.** Quantitative risk assessments of debris flows and other hydrogeological hazards require the analyst to predict damage potentials. A common way to do so is by use of proportional loss functions. In this paper, we analyze a uniquely rich dataset of 132 buildings that were damaged in one of five large debris flow events in Switzerland. Using the double generalized linear model, we estimate proportional loss functions that may be used for various prediction purposes including hazard mapping, landscape planning, and insurance pricing. Unlike earlier analyses, we control for confounding effects of building characteristics, site specifics, and process intensities as well as for overdispersion in the data. Our results suggest that process intensity parameters are the most meaningful predictors of proportional loss sizes. Cross-validation tests suggest that the mean absolute prediction errors of our models are in the range of 11 %, underpinning the accuracy of the approach.

## 1 Introduction

Landslides are a major hydrogeological hazard in mountain ranges all over the globe. Debris flows are a particularly destructive form of fast-moving landslides. Over the past 25 yr, debris flows have killed more than 200 people and caused more than €5 billion in economic damages throughout the European Alps (rough estimate based on Guzzetti et al., 2005; Fuchs, 2009; Hilker et al., 2009). In Switzerland alone, debris flows have caused direct economic damages of €340 million, which equals about 5 % of the total amount of losses from hydrological hazards during this period (Hilker et al., 2009).

It is therefore not surprising that considerable effort has been made to understand the physical processes of debris

flows (Iverson, 1997; Major and Pierson, 1992; Rickenmann, 1999; Sosio and Crosta, 2009). Less emphasis has been given to damages caused by debris flows. The quantification of potential losses is, however, a building block of any quantitative risk assessment (Bonachea et al., 2009; Bründl et al., 2009), and serves as the foundation for the planning of mitigation measures (Dai et al., 2002; Romang et al., 2003; Margreth and Romang, 2010).

As the size and force of a debris flow is a priori unknown, quantitative risk assessments are based on predictions of damages that could occur under different hazard scenarios. The term *physical vulnerability* is commonly used to refer to the proportional loss (also known as relative damage) that an element at risk faces as result of a specific hazard impact (Fuchs, 2009; Uzielli et al., 2008; Hollenstein, 2005). This impact is determined by two factors. One is the intensity of the hazard, which may be defined as “a set of spatially distributed parameters describing the destructiveness” of the hazard process (Hungr, 1997). The other is the susceptibility of the element at risk, which describes the propensity of a building or other infrastructure to suffer damage from a specific hazard impact (Uzielli et al., 2008).

Empirical analyses that relate damages from debris flows and other landslide processes to intensity and susceptibility factors are still rare; see Totschnig and Fuchs (2013) for a recent review of the literature. Most existing studies use either descriptive methods (e.g., Bell and Glade, 2004) or ordinary least-square (OLS) regression (e.g., Fuchs et al., 2007) to explain the proportional loss  $y = d/v$ , defined as the ratio of observed damage  $d$  to total value  $v$ .

While of practical value, there are some drawbacks with these conventional approaches. The descriptive approach does not make full use of the available information and allows only for qualitative damage predictions. OLS regression

can incorporate intensity and susceptibility factors, but its applicability is limited by the fact that regression estimates are unconstrained, allowing for loss estimates to take on implausibly large or small values.

Since the proportional loss is by construction a relative measure, valid models must provide estimates of  $y$  restricted to the  $[0, 1]$  interval. One can constrain OLS to this range, but the resulting censored linear model bears some unattractive features (Kieschnick and McCullough, 2003; Fox, 2008). It is inherently unstable because it critically depends on the values at which minimal and maximal loss is predicted, it is inexpedient as soon as more than one predictor variable is included, the abrupt truncation of the linear slope of  $y$  is unrealistic as one would expect a smooth rather than a kinked relationship between the proportional loss and the predictor variables, and its error variance is most likely not constant (i.e., heteroscedastic) because regressions involving data from the unit interval typically display more variation around the mean than at the limits of the  $[0, 1]$  interval.

Generalized linear models (GLMs) provide a natural framework for the estimation of proportional data (De Jong and Heller, 2008). Geoscientists have previously made use of this property to predict the occurrence probability of debris flow events (Ohlmacher and Davis, 2003; Sepulveda and Padilla, 2008; Griffiths et al., 2004). Yet apart from recent studies by Totschnig et al. (2011) and Totschnig and Fuchs (2013), GLMs have to the best of our knowledge not been used to analyze proportional loss data of debris flow events.

Moreover, there is a caveat to the application of standard GLMs such as logistic or probit regression to proportional loss data. GLMs typically assume that the error variance is constant across observations. This assumption is likely to be violated whenever damages are spatially or temporally correlated. Ignoring the non-constant nature of the error variances may result in a misrepresentation of the true dispersion and erroneous inferences about the parameters of interest (De Jong and Heller, 2008).

In this paper we use the double generalized linear model (DGLM) introduced by Smyth (1989) to explore a unique dataset comprising insurance values of 132 buildings. These buildings were damaged by five debris flow events, which occurred in Switzerland during the late 1980s and early 2000s. While the DGLM provides the same flexibility as standard GLMs, it adjusts for the overdispersion typically observed in proportional losses, and hence lends itself for the statistical analysis of our dataset.

The remainder of the paper is structured as follows. Section 2 describes our empirical strategy, our dataset, and the statistical approach in more detail. Section 3 presents regression results and sensitivity tests that assess the prediction accuracy of the estimated proportional loss functions. In Sect. 4, we sum up with some remarks on the usefulness of our approach for the practical management of debris flow hazards.

## 2 Data and methods

### 2.1 Conceptual framework

The conceptual framework of our empirical analysis of physical vulnerability toward debris flow events is illustrated in Fig. 1. Based on Dai et al. (2002), we identified three building blocks for the statistical analysis of debris flow risk: (1) the runout area, (2) the process intensity, and (3) the elements at risk. In the empirical analysis, we seek to explain observed proportional losses by predictor variables characterizing the three building blocks.

In the empirical analysis, we define the runout area by its roughness and its slope steepness. We characterize the process intensity by the impact angle between the tangent to the flow path of the debris flow at the point of impact and the plane tangent to the struck surface at the point of impact, the flow depth and velocity, and their interaction, which is widely accepted as an intensity indicator (Kreibich et al., 2009). Flow depth and velocity are used not only as criteria for distinguishing intensity zones for hazard mapping of debris flows (Loat and Petrascheck, 1997) but also in risk analyses in Switzerland (Bründl et al., 2009), which was the main argument for preferring them to other metrics such as, e.g., the velocity head used only for comparison reasons in this paper (see Sect. 2.2). Elements at risk are described by their type (concrete, lightweight, or mixed construction), purpose (agricultural, industrial, or residential use), and the presence or absence of object-specific protection measures such as reinforced walls or sheet pilings. Below, the data collection process is described in more detail.

### 2.2 Data

We focused on five major debris flow events that had occurred in the Swiss Alps during the late 1980s and early 2000s. According to the size classification by Jakob (2005) these events all belonged to class 4 debris flows with total volumes ranging from 10 000 to 75 000 m<sup>3</sup>. They caused damages with a total amount of €25 million on 455 buildings corresponding to average damages of roughly €50 000 per affected building. Of these 455 buildings, 132 were eventually selected for analysis. Data selection was based on the availability of sufficiently accurate documentation of the runout area, the process intensity, and the elements at risk (including their total values and the observed damages). Damage observations were split approximately evenly across hazard sites. Figure 2 details out the geographical distribution of the debris flow events and the analyzed damages.

In recent years debris flows in Switzerland have been registered in a central database (Hilker et al., 2009). However, the event descriptions in this database are insufficient for object-specific statistical analyses. Therefore, we had to improve the data quality (see Romang, 2004; Kimmerle, 2002). We assessed type and purpose of each of the buildings based

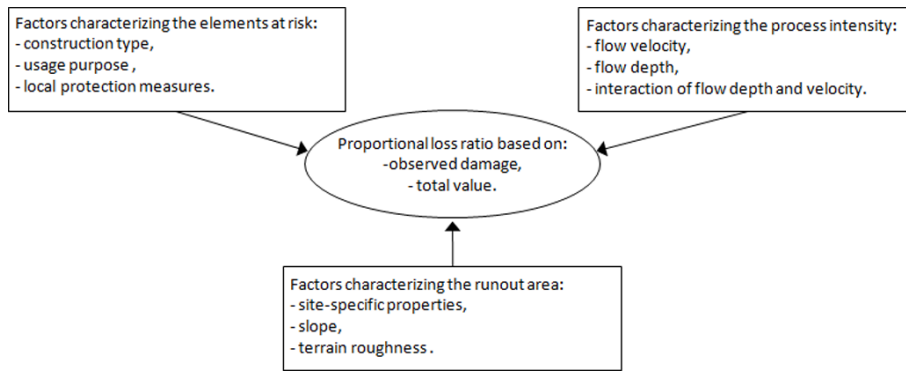


Fig. 1. Conceptual framework for the empirical analysis of proportional loss data.

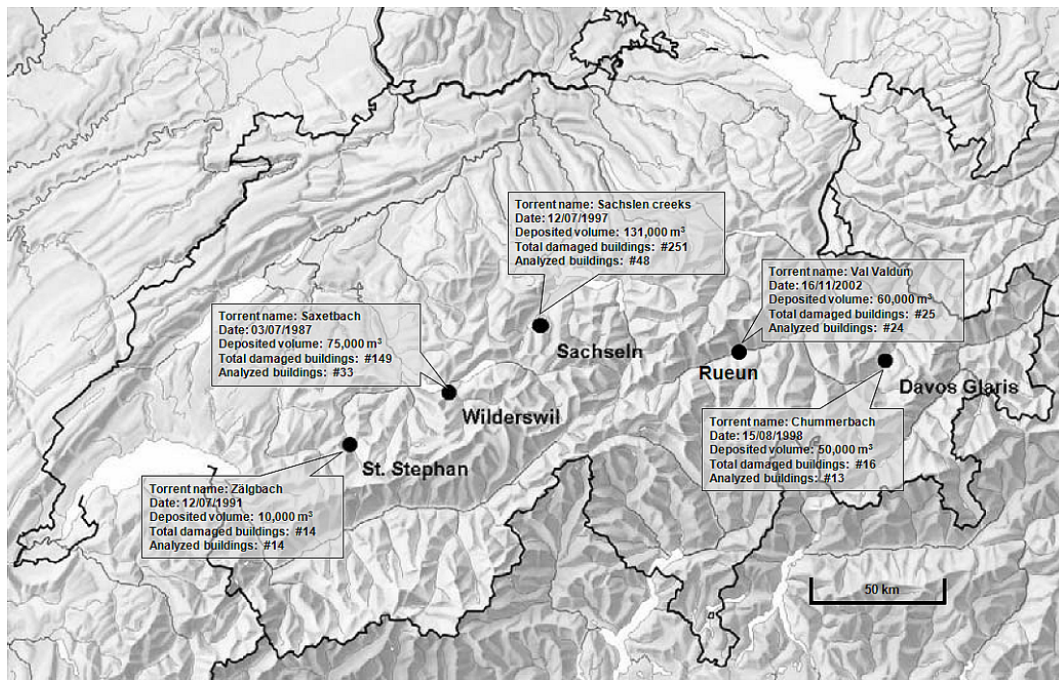


Fig. 2. Summary of the analyzed hazard sites including information about the location, date, total debris volume, and number of affected and of analyzed buildings. ©Federal Office of Topography swisstopo.

on field observations and interviews with affected property owners and other witnesses of the debris flow events. Absolute and relative damages were ascertained based on insurance reports of actuarial building values and the costs of repair or replacement. The process intensity of the analyzed debris flow events was reconstructed from back modeling with FLO-2D (O’Brien et al., 1993) and back calculations of the flow velocity, flow depth, and impact pressures (Hungri et al., 1984). Information about the runout topography, roughness, and inclination was collected directly in the field. Table 1 gives a descriptive summary of the compiled dataset.

Data quality is accurate for all parameters that could be directly observed or measured. The intensity parameters could, however, not be measured directly and their quantification is

therefore fraught with some uncertainty. We reduced this uncertainty by combining back calculations, engineering rules of thumb, and eyewitness reports (Romang, 2004). In particular, we could reconstruct in many cases the maximum deposition height ( $H_{\max D}$ ) based on marks left on exterior walls. According to the Bernoulli principle,  $H_{\max D}$  equals the velocity head  $V^2/(2 \cdot g)$  (with  $g = 9.81 \text{ m s}^{-2}$ ). By solving the equation for  $V = \sqrt{2 \cdot g \cdot H_{\max D}}$  we obtained an estimate of the flow velocity, which we then compared to the flow velocity estimated based on the procedure described in Egli (2005). If the discrepancy between both estimates was larger than  $1 \text{ m s}^{-1}$ , we reassessed the flow velocity.

The outlined data validation process significantly improved the data quality. Yet we emphasize that in the

**Table 1.** Descriptive statistics of the compiled dataset ( $N = 132$ ).

Variables	Mean/ Frequency [SD]	Min	Max
<b>Hazard site</b>			
Saxetbach [-]	0.25	0	1
Zälgbach [-]	0.36	0	1
Sachseln creeks[-]	0.11	0	1
Chummerbach [-]	0.10	0	1
Val Valdun [-]	0.18	0	1
<b>Runout roughness</b>			
Open land [-]	0.38	0	1
Built-over land [-]	0.25	0	1
Forest land [-]	0.05	0	1
Road [-]	0.32	0	1
<b>Runout slope</b>			
< 2 % inclination [-]	0.10	0	1
2–5 % inclination [-]	0.31	0	1
5–10 % inclination [-]	0.36	0	1
More than 10 % inclination [-]	0.23	0	1
<b>Construction type</b>			
Lightweight construction [-]	0.21	0	1
Mixed construction [-]	0.60	0	1
Concrete construction [-]	0.19	0	1
<b>Usage type</b>			
Residential building [-]	0.65	0	1
Industrial building [-]	0.24	0	1
Agricultural building [-]	0.11	0	1
<b>Building properties</b>			
Total value [k€]	617 [1517]	2.67	16 645
Damage [k€]	41 [92]	0.19	906
Proportional loss ratio [-]	0.15 [0.24]	$7 \times 10^{-4}$	1
Local protection measures [-]	0.21	0	1
Impact angle [°]	62 [35]	0	90
<b>Process intensity</b>			
Flow depth [cm]	64 [56]	10	250
Flow velocity [cm s <sup>-1</sup> ]	239 [80]	100	500

SD is the standard deviation.

aftermath of an event a clear distinction between debris flows and hyper-concentrated flows is not always possible because different hazard processes may coincide. For example, the runout area might first be hit by a debris flow surge and subsequently flooded and eroded again (Pierson, 2005). We are unable to differentiate between such transitional forms of flowing and flooding processes. For the purpose of our analysis, this limitation seems acceptable. Totschnig and Fuchs (2013) provide further support for this assumption.

### 2.3 Statistical approach

The statistical modeling of proportional loss functions in the standard GLM framework is based upon three components (Fox, 2008): (1) the loss function,  $\eta_i = \mathbf{X}_i \times \boldsymbol{\beta}$ , where  $\mathbf{X}_i$  is a linear combination of predictor variables for object  $i$ , and  $\boldsymbol{\beta}$  is the corresponding vector of coefficients; (2) a probability distribution of proportional loss from the exponential

family; and (3) an invertible link function  $g$  that maps the expected proportional loss  $E[y_i] = \mu_i$  onto the predictors so that  $g(\mu_i) = \eta_i$ .

Specific link functions have emerged for the analysis of proportional loss data (De Jong and Heller, 2008). The canonical link is the logit

$$\eta_i = \log(d_i/(v_i - d_i)) = \log(y_i/(1 - y_i)),$$

where  $v_i$  and  $d_i$  denote the total value of and the damage to object  $i$  both measured in units of k€. Since the log odds ratio approaches negative infinity as  $y_i$  approaches 0, and moves toward positive infinity as  $y_i$  approaches 1, loss predictions are restricted to the [0, 1] interval.

The logit link gives rise to the standard logistic regression model

$$y_i = (1 + \exp(-\eta_i))^{-1},$$

which can be evaluated by maximum likelihood methods. Applied to polytomous data, the model assumes that the proportional loss  $y_i$  observed at object  $i$  is binomially distributed:  $y_i \sim B(v_i, d_i) \forall i \in N$ , where  $N$  is the set of all observations. This implies that each object at risk can be divided into  $n_i = v_i/\text{k€}$  equal units, which the model treats as if they were independent Bernoulli trials with probability  $p_i = d_i/v_i$ .

This distributional assumption becomes unrealistic when the damage potential is spatially or temporally correlated. A simple example is a debris flow hitting a residential building of total value  $v_i$ , but only the ground floor is affected. Now, assume that the first floor bears less (or more) valuables than the ground floor. In this case, the dispersion across the observed units of proportional loss is larger than warranted by the constant error variance assumption of the logistic regression model:  $\text{Var}[y_i] = \phi y_i(1 - y_i)$ , where  $\phi$  is a dispersion parameter common to all observations.

The DGLM generalizes the above error variance by inclusion of an observation-specific dispersion parameter  $\phi_i$ , which is contingent on some predictor variables to account for systematic variations in the error variance (Smyth, 1989). The observation-specific dispersion parameter can be expressed as  $h(\phi_i) = \mathbf{Z}_i \times \boldsymbol{\lambda}$ , where  $\mathbf{Z}_i$  is a vector of predictors and  $\boldsymbol{\lambda}$  is the corresponding coefficient vector. (Note that  $\mathbf{Z}_i$  can be a perfect subset of  $\mathbf{X}_i$ . Identification is warranted as long as  $\boldsymbol{\lambda}$  comprises a freely estimable intercept.)

The joint estimation of the dispersion submodel  $h(\phi_i)$  and the loss submodel  $g(\mu_i)$  increases the precision in the coefficient estimates of interest because observations with larger variation are penalized (i.e., they are down-weighted in the likelihood maximization). Standard errors for the predictors of both the loss and the dispersion submodels are calculated from the inverse Fisher-information matrix and Wald tests can be used to infer the significance level of a specific predictor (Smyth, 1989). The interpretation of the coefficient estimates from the DGLM with logit link is therefore identical

to the interpretation of those obtained from a standard logistic regression model.

A virtue of this class of models is the ease with which coefficient estimates can be converted to and interpreted as risk ratios (De Jong and Heller, 2008). Risk ratios are particularly insightful to compare the vulnerability of different types of buildings. In general terms, the risk ratio is defined as

$$RR[x'] = \frac{E[y|x' = 1, \bar{X}']}{E[y|x' = 0, \bar{X}']} = \frac{1 + \exp(-\bar{X}'\beta)}{1 + \exp(-(\bar{X}'\beta + \beta_{x'}))}, \quad (1)$$

where  $\beta_{x'}$  is the coefficient on the characteristic of interest  $x'$ .  $E[y|\bar{X}', x' = 1]$  and  $E[y|\bar{X}', x' = 0]$  denote the expected proportional loss for buildings that do ( $x' = 1$ ) and do not have ( $x' = 0$ ) this characteristic, keeping all other predictors  $\bar{X}'$  fixed at sample means.

A common assumption is that the log-transformed risk ratio is normally distributed:  $\log(RR[x']) \sim N(\mu, \sigma^2)$ . The mean is equal to the log of the expected risk ratio:  $\mu = \log(E[y|x' = 1, \bar{X}'] / E[y|x' = 0, \bar{X}'])$ . Using the delta method (Fox, 2008), the variance can be approximated by

$$\sigma^2 \approx \frac{\text{Var}[y|x' = 1, \bar{X}']}{E[y|x' = 1, \bar{X}']^2} + \frac{\text{Var}[y|x' = 0, \bar{X}']}{E[y|x' = 0, \bar{X}']^2} - 2 \cdot \frac{\text{Cov}[y|x' = 1, \bar{X}'; y|x' = 0, \bar{X}']}{E[y|x' = 1, \bar{X}'] E[y|x' = 0, \bar{X}']}, \quad (2)$$

where  $\text{Var}[\cdot]$  and  $\text{Cov}[\cdot, \cdot]$  denote the sample variance and covariance. (A detailed derivation of Eq. 2 is given in Appendix A1.) Based on this, the 95 % confidence interval (95 % CI) around the risk ratio is calculated as  $\text{CI} = \exp[\mu \pm 1.96 \cdot \sigma]$ .

### 3 Results

#### 3.1 Regression models

We present coefficient estimates for three specifications of the DGLM with logit link function. These specifications differ with respect to the predictor variables included in the loss and dispersion functions.

##### 3.1.1 Model I

The first model specification explains observed proportional losses solely through intensity predictors. The dispersion function includes a global constant and site-specific constants that control for differences in dispersion across debris flow sites. The loss function includes the impact angle ( $A$ ), flow depth ( $D$ ), flow velocity ( $V$ ), and their interaction ( $D \cdot V$ ), which we take as an intensity indicator. We use log transformations to reduce skew in the distribution of these predictors. Note that  $\log D \cdot \log V$  is a non-linear transformation of  $\log D \cdot V^2$ , which others have proposed as a surrogate

measure of impact force (Jakob et al., 2012; Kreibich et al., 2009).

Regression results reported in Table 2 indicate that the expected loss is significantly determined by the four intensity predictors. The coefficient on the logged impact angle is positive, but the effect of a larger impact angle on the loss function is negligibly small. To see this, consider two impact angles  $A_1$  and  $A_2 = 2 \cdot A_1$ . Since the predictor is log-transformed, the effect of doubling the impact angle while keeping everything else constant is  $0.224 \cdot (\log A_2 - \log A_1) = 0.224 \cdot \log A_2 / A_1 \approx 0.155$ . In other words, the loss function increases by only 0.155 units for every doubling of the impact angle.

Coefficients on the flow depth and velocity are both negative, but the coefficient on the interaction term is positive. This implies that proportional losses grow with increasing flow depth ( $\partial y / \partial D > 0$ ) and velocity ( $\partial y / \partial V > 0$ ), but the growth rate is marginally decreasing ( $\partial^2 y / \partial D^2 < 0$ ,  $\partial^2 y / \partial V^2 < 0$ ). In Fig. 3, we visualize this relationship by plotting the proportional losses predicted from Model I for debris flow depths and velocities typically assumed in the design of local protection strategies (Egli, 2005). The reader is warranted that  $D$  and  $V$  are positively correlated (Pearson's  $\rho = 0.26$ ,  $P < 0.01$ ), so that combinations with low (high) velocities and large (small) depths are unlikely to be observed in reality. Therefore, the interpretation of model predictions for such out-of-bounds combinations is meaningless.

The significant intercept of the dispersion function underlines that there is overdispersion in our data. Moreover, the dispersion varies across hazard sites and is significantly lower for the Chummerbach event, suggesting more homogeneous damages at this hazard site.

##### 3.1.2 Model II

The second model specification adds site-specific and object-specific attributes to the loss function. A likelihood-ratio test confirms that the model fit is significantly improved by the inclusion of the additional predictors ( $\chi^2_{9\text{df}}: 15.8$ ,  $P < 0.05$ ). No differences are observed across the five hazard sites. With the exception of the insignificant coefficient on the impact angle, coefficient estimates on the intensity predictors are similar to those obtained with Model I. This suggests site-specific factors in Model II do not have a strong confounding relationship with the process intensity but help to explain variability in the data.

Coefficient estimates on object-specific characteristics imply that proportional losses at industrial buildings are significantly larger than those observed for residential buildings, while proportional losses at agricultural buildings are statistically not different from those observed for residential buildings. Concrete buildings are significantly less vulnerable to debris flow damages, as are buildings that had local protection measures installed.

**Table 2.** Regression results of different DGLM specifications; standard errors are reported in brackets.

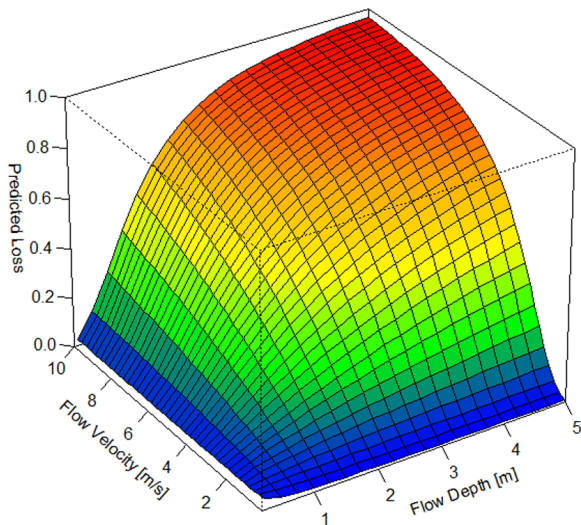
	Model I <sup>a</sup>	Model II <sup>a</sup>	Model III <sup>a</sup>	
Loss function	Intercept	8.001 (7.99)	10.784 (8.89)	15.021 (9.53)
	Hazard site (default: Saxetbach)			
	Zälgbach	–	–0.096 (0.30)	–0.467** (0.24)
	Sachseln	–	–0.111 (0.36)	–0.126 (0.38)
	Chummerbach	–	–0.046 (0.44)	0.194 (0.32)
	Val Valdun	–	0.412 (0.38)	–0.446 (0.38)
	Runout roughness (default: open land)			
	Built-over land	–	–	0.351 (0.29)
	Forest land	–	–	0.239 (0.65)
	Road	–	–	–0.187 (0.23)
	Runout slope (default: less than 2 % inclination)			
	2–5 % inclination	–	–	1.322*** (0.41)
	5–10 % inclination	–	–	1.452*** (0.43)
	More than 10 % inclination	–	–	1.181** (0.46)
	Usage type (default: residential building)			
	Agricultural building	–	0.407 (0.51)	1.214*** (0.43)
	Industrial building	–	0.395* (0.24)	0.380* (0.22)
	Construction type (default: mixed construction)			
	Lightweight construction	–	–0.070 (0.34)	–0.204 (0.36)
	Concrete construction	–	–0.669** (0.31)	–0.639*** (0.24)
Building properties				
Local protection measures	–	–0.608* (0.33)	–0.742*** (0.25)	
Log of impact angle	0.224** (0.09)	0.018 (0.09)	0.099 (0.07)	
Process intensity				
Log of flow depth [cm]	–3.579* (1.93)	–4.342** (2.12)	–5.131** (2.20)	
Log of flow velocity [ $\text{cm s}^{-1}$ ]	–2.871** (1.47)	–3.091* (1.61)	–4.257** (1.74)	
Log of flow velocity $\times$ log of flow depth [ $\text{cm}^2 \text{s}^{-1}$ ]	0.853* (0.35)	0.960** (0.38)	1.124*** (0.41)	
Dispersion function	Intercept	3.980*** (0.25)	3.840*** (0.25)	57.769*** (10.51)
	Hazard site (default: Saxetbach)			
	Zälgbach	–0.272 (0.32)	–0.125 (0.32)	–0.064 (0.33)
	Sachseln	–0.667 (0.45)	–1.068** (0.45)	0.127 (0.45)
	Chummerbach	–1.236*** (0.46)	–0.953** (0.46)	–0.987** (0.50)
	Val Valdun	0.049 (0.38)	–0.060 (0.38)	0.154 (0.41)
	Process intensity			
	Log of flow depth [cm]			–13.010*** (2.48)
Log of flow velocity [ $\text{cm s}^{-1}$ ]			–10.463*** (1.94)	
Log of flow velocity $\times$ log of flow depth [ $\text{cm}^2 \text{s}^{-1}$ ]			2.500*** (0.46)	
Model properties	Log likelihood at convergence	–593.3	–585.4	–557.4
	McFadden's pseudo- $R^2$	0.08	0.10	0.14
	Mean absolute prediction error <sup>b</sup>	0.109	0.109	0.110

<sup>a</sup> Significance coding for Wald tests on coefficient estimates: \*\*\* < 0.01, \*\* < 0.05, \* < 0.1; <sup>b</sup> see the description in Sect. 3.2.

With Eq. (1) we can study how much the construction type, the usage purpose, or the implementation of local protection measures reduces or increases the expected proportional loss. As an illustrative example, consider two buildings of the same type and value but one is protected by sheet pilings, while the other is not. The risk ratio is  $\text{RR}[x' = \text{protection}] = 0.62(\text{CI}: 0.42, 0.89)$ , meaning that we expect a  $\sim 38\%$  smaller loss at the protected building. Similar risk ratios for other predictors of interest are reported in Table 3.

### 3.1.3 Model III

This specification includes all available information in the loss function. Additionally, we amend the dispersion function with predictors of the flow height, flow velocity, and their interaction. In terms of statistical fit, the improvement over Model II is significant ( $\chi^2_{10\text{df}}: 55.9, P < 0.001$ ). This improvement is largely due to the improved capture of heterogeneity in the dispersion function. Inclusion of the intensity predictors into the dispersion function accounts for



**Fig. 3.** Proportional loss predictions by Model I for commonly observed debris flow depths and velocities. To construct this plot, the impact angle was fixed at the sample means of  $64^\circ$ .

variation in dispersion at the building level rather than the hazard site level.

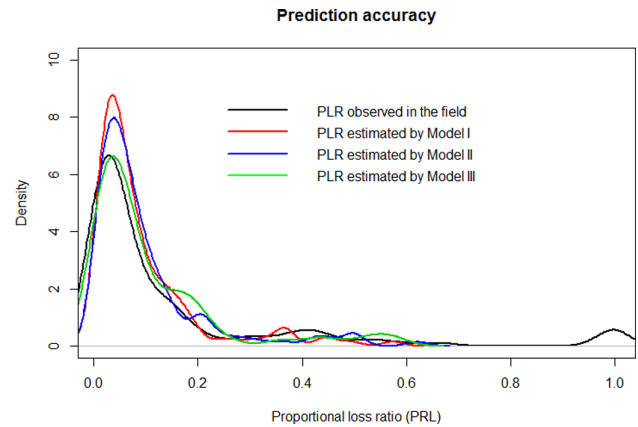
The site-specific constants indicate that losses observed at the Zalgbach site are significantly smaller than those observed at the other hazard sites. Once we control for event-specific factors, the other sites are statistically not different from each other. Information about the runout cover is not predictive. However, proportional losses increase with runout steepness.

Results of Model III support the finding that industrial buildings are more vulnerable to debris flow damages than residential buildings. Controlling for additional dispersion in the data, we now also find a significant effect on agricultural buildings. Concrete buildings and buildings where local protection measures were installed are significantly less vulnerable to debris flows than mixed or light constructions (Table 3). Coefficient estimates on the intensity predictors are again similar to those obtained with Models I and II, confirming that they are the most important loss predictors.

### 3.2 Prediction accuracy

We use leave-one-out cross validation (LOOCV) to assess the out-of-sample prediction accuracy of our models (Wilks, 2006). This resampling procedure repeatedly fits the DGLM using  $N - 1$  observations as training set to make a prediction for the left-out observation. Then, the predicted loss for this observation is compared to the observed loss and prediction accuracy is measured in terms of the mean absolute prediction error across every observation.

The mean absolute prediction errors for Models I–III are 0.108, 0.109, and 0.110, respectively (Table 2). This suggests that loss predictions based on the three models are on average



**Fig. 4.** Kernel density plot (Gaussian kernel, bandwidth: 1) of observed versus predicted proportional losses for the 132 damaged buildings.

within a range of  $\pm 11\%$  around the observed loss size and that none of the models outperforms the others in terms of prediction quality. This is, however, not to be confused with the statistical fit, which measures the explanatory power of the included predictors.

To see whether there are ranges within the  $[0, 1]$  interval where the predictions are particularly inaccurate, we used Models I–III to predict the proportional losses on the full dataset, and compared predicted losses with observed losses. Figure 4 illustrates that all three models do a decent job of predicting small to medium losses. However, the models do not predict large losses or even total losses. We conclude that while our approach is useful to predict damages of common sizes, it is less suited to predict extreme damages. By definition, these are rare events and we miss sufficient observations in the extreme range of damages that would allow for us to better calibrate the model.

## 4 Conclusions

The assessment of potential damages is a key ingredient of any risk assessment study. As risk-based decisions play an ever more important role in natural hazard management (Brundl et al., 2009), there has been growing interest in the development of predictive loss functions that could be used to quantify damage potentials for various natural hazards.

In this paper, we have outlined a framework to estimate proportional loss functions for debris flow events. The framework lends itself to the analysis of other hydrological hazards. Unlike standard regression models, the presented DGLM can deal with problems of overdispersion, and is therefore a valuable tool for empirical analyses of natural hazard damages.

We have used the DGLM to analyze what is maybe the richest dataset on debris flow damages to date in Switzerland. We find that flow depth, flow velocity, and their interaction

**Table 3.** Risk ratios for factors characterizing the elements at risk; 95 % confidence intervals are reported in brackets.

	Model II	Model III
Usage type (default: residential building)		
Agricultural building	1.364 (CI: 1.042, 1.787)	2.383 (CI: 1.122, 5.059)
Industrial building	1.349 (CI: 1.045, 1.742)	1.315 (CI: 0.997, 1.734)
Construction type (default: mixed construction)		
Lightweight construction	0.947 (CI: 0.903, 0.994)	0.862 (CI: 0.742, 1.002)
Concrete construction	0.584 (CI: 0.397, 0.859)	0.610 (CI: 0.418, 0.890)
Local protection measures	0.615 (CI: 0.424, 0.894)	0.563 (CI: 0.362, 0.875)

CI is the confidence interval.

are the most meaningful predictors of relative damage. Industrial and agricultural buildings tend to be more vulnerable to debris flows than residential buildings. This is most likely due to different value concentrations of the ground floor. Many agricultural and industrial buildings are single-story buildings, which by nature have an increased damage potential within the reach of debris flow surges. Residential buildings, in contrast, often have their values distributed over several stories so that the damage potential in the ground floor is relatively small compared to the total damage potential. This underlines the importance of controlling for overdispersion at the building level.

We find some insightful results with regard to building characteristics. These are of particular interest because property owners can alter them through self-protection efforts. Using the risk ratios reported in Table 3, the damage to a building without local protection measures such as sheet piling or concrete enforcements is predicted to be 1.6–1.8 times as high as the damage to an otherwise identical non-protected building. Similarly, the damage predicted to a concrete building is only about 60 % of the damage expected to an otherwise identical non-concrete building, suggesting that structural adaptation to the local environment is an effective means to reduce debris flow damages (Margreth and Romang, 2010).

How accurate are the results? As a LOOCV analysis has confirmed, proportional loss predictions generated with the DGLM approach are quite precise, with mean absolute prediction errors in the range of 11 %. From a practical point of view, this prediction quality is more than sufficient as the uncertainty introduced by the loss predictions is far smaller than that of other parameters entering the risk assessment for debris flow (Schaub and Bründl, 2010).

How do the results compare to the proportional loss functions found in recent studies by Totschnig et al. (2011) and Totschnig and Fuchs (2013)? We did some rough calculations comparing the proportional loss function obtained from our Model I with the best-fitting loss functions reported in Eq. (7) of Totschnig et al. (2011) and Eq. (3) of Totschnig and Fuchs (2013), respectively. If we keep flow velocity and impact angle fixed at sample means, we obtain loss predictions that are about 50 % lower than those of Totschnig et al. (2011); if we assume a flow velocity of  $7 \text{ ms}^{-1}$  (the upper

range of the values Egli (2005) proposed for debris flows in flat terrain), loss predictions obtained from both models are within a range of 10 % (see Appendix A2 for a visualization). We conclude that, under reasonable assumptions, our loss predictions are in close agreement with those of Totschnig et al. (2011) and Totschnig and Fuchs (2013).

What is the practical value of the results? We may think of several ways to make use of proportional loss predictions by our approach. Here are but two examples. First, modern GIS techniques allow for combining proportional loss functions with hazard maps and databases of insurance contracts to create proper risk maps. Such risk maps could help hazard managers to identify at risk areas and to prioritize among mitigation needs based on the actual damage potential. Second, private property insurers could use the loss predictions to offer individualized contracts that better reflect each homeowner's risk. This would lead to a reduction in the overall volume of premiums paid by insurance holders because fair pricing reduces "moral hazard" – the tendency of the most-exposed insurance holders not to take self-protection measures, as there are no incentives to do so (Raschky and Weck-Hannemann, 2007). State building insurers (in 19 cantons of Switzerland) could use our results to get an overview as to where house owners should consider building local protection measures.

In conclusion, we believe that the presented approach bears great potential for the prediction of natural hazard damages. Needless to say that, with more data, we will be able to better calibrate the proportional loss functions and to improve their prediction accuracy. It is therefore of great importance to build up and maintain central damage databases – a very good example is the HOWAS 21 database for flood losses (Thieken et al., 2009) – that allow for storing of event-specific information. Our analysis emphasizes that flow depths and heights, observed damages, and total values of damaged buildings are of particular interest.

## Appendix A

### A1 Derivation of Eq. (2) by the delta method

The variance of the log-transformed risk ratio is a non-linear function, and hence has to be approximated. A common method to do so is the delta method; see, e.g., Fox (2008).



Before we derive the variance of  $\log(\text{RR}[x'])$ , we introduce a more compact notation. Define

$$\bar{y}_j \equiv E[y|x' = j, \bar{X}'] \quad \text{with } j = \{0, 1\},$$

$$y_{ij} \equiv y_i|x' = j, \bar{X}' \quad \forall i \in N \quad \text{and } j = \{0, 1\},$$

and

$$f \equiv \log \bar{y}_1 / \bar{y}_0 = \log \bar{y}_1 - \log \bar{y}_0.$$

By the delta method, the variance of the non-linear function  $f$  can be approximated as

$$\text{Var}[\log \text{RR}(x')] = \nabla f \Sigma \nabla f^T,$$

where  $\nabla f$  denotes the gradient vector

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial \bar{y}_1} & \frac{\partial f}{\partial \bar{y}_0} \end{bmatrix} = \begin{bmatrix} \frac{1}{\bar{y}_1} & \frac{1}{\bar{y}_0} \end{bmatrix},$$

and  $\Sigma$  denotes the variance–covariance matrix

$$\Sigma = N^{-1} \begin{bmatrix} \sum_{i=1}^N (y_{i1} - \bar{y}_1)^2 & -\sum_{i=1}^N (y_{i0} - \bar{y}_0)(y_{i1} - \bar{y}_1) \\ -\sum_{i=1}^N (y_{i1} - \bar{y}_1)(y_{i0} - \bar{y}_0) & \sum_{i=1}^N (y_{i0} - \bar{y}_0)^2 \end{bmatrix}.$$

By inserting, we have

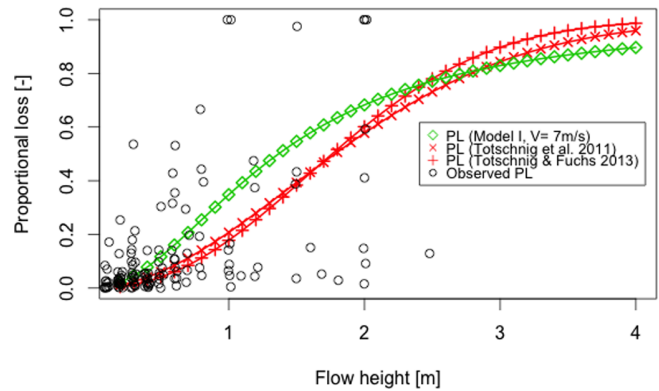
$$\nabla f \Sigma \nabla f^T = \begin{bmatrix} \frac{1}{\bar{y}_1} & \frac{1}{\bar{y}_0} \end{bmatrix} \cdot N^{-1} \begin{bmatrix} \sum_{i=1}^N (y_{i1} - \bar{y}_1)^2 & -\sum_{i=1}^N (y_{i0} - \bar{y}_0)(y_{i1} - \bar{y}_1) \\ -\sum_{i=1}^N (y_{i1} - \bar{y}_1)(y_{i0} - \bar{y}_0) & \sum_{i=1}^N (y_{i0} - \bar{y}_0)^2 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{\bar{y}_1} \\ \frac{1}{\bar{y}_0} \end{bmatrix} = N^{-1} \begin{bmatrix} \frac{\sum_{i=1}^N (y_{i1} - \bar{y}_1)^2}{\bar{y}_1^2} + \frac{\sum_{i=1}^N (y_{i0} - \bar{y}_0)^2}{\bar{y}_0^2} \\ -2 \frac{\sum_{i=1}^N (y_{i1} - \bar{y}_1)(y_{i0} - \bar{y}_0)}{\bar{y}_1 \bar{y}_0} \end{bmatrix} = \frac{\text{Var}[y_{i1}]}{\bar{y}_1^2} + \frac{\text{Var}[y_{i0}]}{\bar{y}_0^2}$$

$$-2 \frac{\text{Cov}[y_{i1}, y_{i0}]}{\bar{y}_1 \bar{y}_0},$$

which is equivalent to Eq. (2) in the main text.

## A2 Visual comparison with other proportional loss functions



**Fig. A1.** Comparison of the proportional loss (PL) function of Model I (evaluated at  $V = 7 \text{ m s}^{-1}$ ) with the best-fitting PL functions of Totschnig et al. (2011: Eq. 7) and Totschnig and Fuchs (2013, Eq. 3). PL is the proportional loss.

*Acknowledgements.* This paper was partly funded through EU FP6 STREP contract no. 081412 (“IRASMOS”). We thank Roland Kimmerle for providing his dataset. We are grateful to Sven Fuchs and three other reviewers, whose comments greatly helped in improving the paper.

Edited by: O. Katz

Reviewed by: S. Fuchs and three anonymous referees

## References

Bell, R. and Glade, T.: Quantitative risk analysis for landslides – Examples from Bífudalur, NW-Iceland, *Nat. Hazards Earth Syst. Sci.*, 4, 117–131, doi:10.5194/nhess-4-117-2004, 2004.

Bonachea, J., Remondo, J., De Terán, J. R. D., González-Díez, A., and Cendrero, A.: Landslide Risk Models for Decision Making, *Risk Anal.*, 29, 1629–1643, doi:10.1111/j.1539-6924.2009.01283.x, 2009.

Bründl, M., Romang, H. E., Bischof, N., and Rheinberger, C. M.: The risk concept and its application in natural hazard risk management in Switzerland, *Nat. Hazards Earth Syst. Sci.*, 9, 801–813, doi:10.5194/nhess-9-801-2009, 2009.

Dai, F. C., Lee, C. F., and Ngai, Y. Y.: Landslide risk assessment and management: an overview, *Eng. Geol.*, 64, 65–87, doi:10.1016/S0013-7952(01)00093-X, 2002.

De Jong, P. and Heller, G. Z.: Generalized Linear Models for Insurance Data, International Series on Actuarial Science, Cambridge University Press, Cambridge, 2008.

Egli, T.: Wegleitung Objektschutz gegen gravitative Naturgefahren, Vereinigung Kantonaler Feuerversicherer, Bern, Switzerland, available at: <http://www.vkf.ch/VKF/Downloads.aspx>, last access: 8 August 2013, 2005 (in German).

Fox, J.: Applied Regression Analysis and Generalized Linear Models, 2 Edn., Sage, Thousand Oaks, CA, 2008.

Fuchs, S.: Susceptibility versus resilience to mountain hazards in Austria – paradigms of vulnerability revisited, *Nat. Hazards Earth Syst. Sci.*, 9, 337–352, doi:10.5194/nhess-9-337-2009, 2009.

- Fuchs, S., Heiss, K., and Hübl, J.: Towards an empirical vulnerability function for use in debris flow risk assessment, *Nat. Hazards Earth Syst. Sci.*, 7, 495–506, doi:10.5194/nhess-7-495-2007, 2007.
- Griffiths, P. G., Webb, R. H., and Melis, T. S.: Frequency and initiation of debris flows in Grand Canyon, Arizona, *J. Geophys. Res.*, 109, F04002, doi:10.1029/2003JF000077, 2004.
- Guzzetti, F., Stark, C. P., and Salvati, P.: Evaluation of flood and landslide risk to the population of Italy, *Environ. Manage.*, 36, 15–36, doi:10.1007/s00267-003-0257-1, 2005.
- Hilker, N., Badoux, A., and Hegg, C.: The Swiss flood and landslide damage database 1972–2007, *Nat. Hazards Earth Syst. Sci.*, 9, 913–925, doi:10.5194/nhess-9-913-2009, 2009.
- Hollenstein, K.: Reconsidering the risk assessment concept: Standardizing the impact description as a building block for vulnerability assessment, *Nat. Hazards Earth Syst. Sci.*, 5, 301–307, doi:10.5194/nhess-5-301-2005, 2005.
- Hungr, O.: Some methods of landslide intensity mapping, in: *Landslide Risk Assessment*, edited by: Cruden, D. and Fell, R., Balkema, Rotterdam, 1997.
- Hungr, O., Morgan, G., and Kellerhals, C.: Quantitative analysis of debris flow torrent hazards for design of remedial measures, *Can. Geotech. J.*, 21, 663–677, doi:10.1139/t84-073, 1984.
- Iverson, R. M.: The physics of debris flows, *Rev. Geophys.*, 35, 245–296, doi:10.1029/97RG00426, 1997.
- Jakob, M.: A size classification for debris flows, *Eng. Geol.*, 79, 151–161, doi:10.1016/j.enggeo.2005.01.006, 2005.
- Jakob, M., Stein, D., and Ulmi, M.: Vulnerability of buildings to debris flow impact, *Nat. Hazards*, 60, 241–261, doi:10.1007/s11069-011-0007-2, 2012.
- Kieschnick, R. and McCullough, B. D.: Regression analysis of variates observed on (0,1): percentages, proportions and fractions, *Stat. Model.*, 3, 193–213, doi:10.1191/1471082X03st053oa, 2003.
- Kimmerle, R.: Schadenempfindlichkeit von Gebäuden gegenüber Wildbachgefahren, Master thesis, University of Bern, Switzerland, 2002 (in German).
- Kreibich, H., Piroth, K., Seifert, I., Maiwald, H., Kunert, U., Schwarz, J., Merz, B., and Thielen, A. H.: Is flow velocity a significant parameter in flood damage modelling?, *Nat. Hazards Earth Syst. Sci.*, 9, 1679–1692, doi:10.5194/nhess-9-1679-2009, 2009.
- Loat, R. and Petrascheck, A.: Consideration of Flood Hazards for Activities with Spatial Impact – Recommendations, Federal Office for Water Management (BWW), Federal Office for Spatial Planning (BRP), Federal Office for the Environment, Forests and Landscape (BUWAL), Bern, available at: [www.bafu.admin.ch](http://www.bafu.admin.ch), last access: 25 June 2013, 1997.
- Major, J. J. and Pierson, T. C.: Debris flow rheology: Experimental analysis of fine-grained slurries, *Water Resour. Res.*, 28, 841–857, doi:10.1029/91WR02834, 1992.
- Margreth, S. and Romang, H. E.: Effectiveness of mitigation measures against natural hazards, *Cold Reg. Sci. Technol.*, 64, 199–207, doi:10.1016/j.coldregions.2010.04.013, 2010.
- O'Brien, J. S., Julien, P. Y., and Fullerton, W. T.: 2-dimensional water flood and mudflow simulation, *J. Hydraul. Eng.-ASCE*, 119, 244–261, 1993.
- Ohlmacher, G. C. and Davis, J. C.: Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA, *Eng. Geol.*, 69, 331–343, 2003.
- Pierson, T. C.: Hyperconcentrated flow – transitional process between water flow and debris flow, in: *Debris Flow Hazards and Related Phenomena*, edited by: Jakob, M. and Hungr, O., Springer, Heidelberg, 2005.
- Raschky, P. A. and Weck-Hannemann, H.: Charity hazard – A real hazard to natural disaster insurance?, *Environmental Hazards*, 7, 321–329, doi:10.1016/j.envhaz.2007.09.002, 2007.
- Rickenmann, D.: Empirical relationships for debris flows, *Nat. Hazards*, 19, 47–77, doi:10.1023/A:1008064220727, 1999.
- Romang, H. E.: *Wirksamkeit und Kosten von Wildbach-Schutzmassnahmen*, Verlag des Geographischen Instituts der Universität Bern, Switzerland, 2004 (in German).
- Romang, H. E., Kienholz, H., Kimmerle, R., and Böll, A.: Control structures, vulnerability, cost-effectiveness: a contribution to the management of risks from debris torrents, in: *Debris-flow hazards mitigation: mechanics, prediction, and assessment*, edited by: Rickenmann, D. and Chen, C., Millpress, Rotterdam, 1303–1313, 2003.
- Schaub, Y. and Bründl, M.: Zur Sensitivität der Risikoberechnung und Massnahmenbewertung von Naturgefahren, *Schweizerische Zeitschrift für Forstwesen*, 161, 27–35, doi:10.3188/szf.2010.0027, 2010 (in German).
- Sepulveda, S. and Padilla, C.: Rain-induced debris and mudflow triggering factors assessment in the Santiago cordilleran foothills, Central Chile, *Nat. Hazards*, 47, 201–215, doi:10.1007/s11069-007-9210-6, 2008.
- Smyth, G. K.: Generalized Linear Models with Varying Dispersion, *J. Roy. Stat. Soc. B Met.*, 51, 47–60, 1989.
- Sosio, R. and Crosta, G. B.: Rheology of concentrated granular suspensions and possible implications for debris flow modeling, *Water Resour. Res.*, 45, W03412, doi:10.1029/2008WR006920, 2009.
- Thielen, A. H., Seifert, I., Elmer, F., Maiwald, H., Haubrock, S., Schwarz, J., Müller, M., and Seifert, J.-O.: Standardisierte Erfassung und Bewertung von Hochwasserschäden, *Hydrol. Wasserbewirts.*, 53, 198–207, 2009 (in German).
- Totschnig, R. and Fuchs, S.: Mountain torrents: Quantifying vulnerability and assessing uncertainties, *Eng. Geol.*, 155, 31–44, doi:10.1016/j.enggeo.2012.12.019, 2013.
- Totschnig, R., Sedlacek, W., and Fuchs, S.: A quantitative vulnerability function for fluvial sediment transport, *Nat. Hazards*, 58, 681–703, doi:10.1007/s11069-010-9623-5, 2011.
- Uzielli, M., Nadim, F., Lacasse, S., and Kaynia, A. M.: A conceptual framework for quantitative estimation of physical vulnerability to landslides, *Eng. Geol.*, 102, 251–256, doi:10.1016/j.enggeo.2008.03.011, 2008.
- Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences*, Elsevier, Amsterdam, 2006.