RESEARCH ARTICLE

10.1002/2017WR021044

Key Points:

- Precipitation and landslide databases were used to define triggering thresholds using ROC formalism
- Thresholds were localized according to erodibility and mean daily precipitation
- Validation shows a robust predictive performance of the rainfall threshold model

Correspondence to:

E. Leonarduzzi,
leonarduzzi@ifu.baug.ethz.ch

Citation:

Leonarduzzi, E., P. Molnar, and B. W. McARDell (2017), Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data, *Water Resour. Res.*, 53, 6612–6625, doi:10.1002/2017WR021044.

Received 3 MAY 2017

Accepted 11 JUL 2017

Accepted article online 15 JUL 2017

Published online 4 AUG 2017

Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data

Elena Leonarduzzi^{1,2} , Peter Molnar¹ , and Brian W. McARDell²
¹Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland, ²Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

Abstract A high-resolution gridded daily precipitation data set was combined with a landslide inventory containing over 2000 events in the period 1972–2012 to analyze rainfall thresholds which lead to landsliding in Switzerland. We colocated triggering rainfall to landslides, developed distributions of triggering and nontriggering rainfall event properties, and determined rainfall thresholds and intensity-duration ID curves and validated their performance. The best predictive performance was obtained by the intensity-duration ID threshold curve, followed by peak daily intensity $I_{\max}$ and mean event intensity I_{mean} . Event duration by itself had very low predictive power. A single country-wide threshold of $I_{\max} = 28 \text{ mm/d}$ was extended into space by regionalization based on surface erodibility and local climate (mean daily precipitation). It was found that wetter local climate and lower erodibility led to significantly higher rainfall thresholds required to trigger landslides. However, we showed that the improvement in model performance due to regionalization was marginal and much lower than what can be achieved by having a high-quality landslide database. Reference cases in which the landslide locations and timing were randomized and the landslide sample size was reduced showed the sensitivity of the $I_{\max}$ rainfall threshold model. Jack-knife and cross-validation experiments demonstrated that the model was robust. The results reported here highlight the potential of using rainfall ID threshold curves and rainfall threshold values for predicting the occurrence of landslides on a country or regional scale with possible applications in landslide warning systems, even with daily data.

1. Introduction

Landslides belong to one of the most impacting natural hazards in alpine landscapes with a high frequency of occurrence and significant economic losses. In Switzerland landslide-related damages were estimated to exceed 0.5 billion USD in the period 1972–2007 [Hilker *et al.*, 2009] and similar levels of economic losses were reported for other alpine countries [Kjekstad and Highland, 2009; Salvati *et al.*, 2010; Trezzini *et al.*, 2013; Klose, 2015]. In the U.S. landslides are estimated to lead to about 2 billion USD damage on average every year [Schuster, 1996]. Therefore, being able to predict landslide occurrence in time and space is important for developing risk maps and early warning systems which would reduce economic losses and fatalities [e.g., Staehli *et al.*, 2015]. The scientific and application challenges to landslide prediction originate in the complex soil processes which determine soil structure and slope stability (predisposing factors) and the environmental conditions which lead to a decrease in soil shear strength and ultimately slope failure (triggering factors).

Rainfall is the most significant triggering factor for shallow hillslope landsliding in an alpine environment, and many empirical and probabilistic approaches to quantify the joint occurrence of rainfall and landslides exist. The most popular approach is the definition of a rainfall intensity-duration (ID) threshold curve which quantifies the landslide triggering rainfall condition. This notion was first formalized in Stevenson [1977] and Caine [1980] and has since been used in many studies to elaborate ID threshold curves for different regions worldwide (see reviews in Guzzetti *et al.* [2007, 2008]). Antecedent rainfall which may increase the susceptibility to landsliding can be included in ID curves [e.g., Glade *et al.*, 2000; Martelloni *et al.*, 2012] or even actual water balance estimates of soil wetness [e.g., Crozier, 1999; Glade, 2000; Schmidt *et al.*, 2008; Ponziani *et al.*, 2012]. There are three typical problems which are encountered in rainfall triggering analyses which make generalizations between studies and the practical application of ID curves for landslide warning systems difficult.

First is the methodological incoherence in defining ID threshold values, e.g., as lower envelopes below which no landslides occur, upper envelopes above which landslides always occur, or intermediate curves

defined by statistical methods. Examples range from simple visual fits [e.g., *Caine*, 1980] to methods based on Bayesian inference to define a minimum threshold [e.g., *Guzzetti et al.*, 2007, 2008; *Brunetti et al.*, 2010] and frequentist approaches which associate thresholds with different exceedance probabilities [e.g., *Brunetti et al.*, 2010; *Meyer et al.*, 2012; *Peruccacci et al.*, 2012; *Brunetti et al.*, 2013; *Gariano et al.*, 2015; *Piciullo et al.*, 2016]. Some studies are only based on triggering events, i.e., rainfall associated with landsliding, while others take into account also nontriggering events by Bayesian approaches [e.g., *Berti et al.*, 2012] or with statistical methods which maximize the probability of detection and minimize false alarms [e.g., *Segoni et al.*, 2009; *Staley et al.*, 2013; *Corsini and Mulas*, 2016]. Furthermore, precipitation event characteristics that may be important for landslide triggering are many, e.g., total rainfall amount, peak intensity, event duration, etc. Therefore, using nontriggering events and a range of precipitation characteristics is important, as this allows for a proper inclusion of the inherent stochasticity in landslide triggering and a level of objectivity in defining rainfall and ID thresholds.

Second is the lack of consistency in landslide and rainfall data. The assumption behind rainfall data in ID curves is that they represent precipitation properties at the site of a landslide, while in practice usually the closest raingauge is used (sometimes many kilometers away). Given the large spatial variability in precipitation in alpine terrain, this colocation problem may lead to large discrepancies in triggering rainfall at the landslide site. Although this problem is acknowledged in most studies, it has been directly addressed in only very few recent works [*Nikolopoulos et al.*, 2014, 2015; *Marra et al.*, 2016] and its effect on published ID curves is still largely unexplored and unknown. In fact, *Gariano et al.* [2015] has shown that adding uncertainty to triggering rainfall, e.g., as a result of the colocation problem, may dramatically affect the predictive performance of ID threshold curve models. Another inconsistency in ID curves is that peak event intensity is commonly used instead of the real triggering intensity just before the landslide occurrence. The latter is often unknown due to inaccurate timing of landslides in the databases [*Staley et al.*, 2013]. Additional uncertainties arise because landslide databases are often limited to large storms or only a few years of records, and do not cover a long enough period which would be representative for the statistics of rainfall events from the climatological point of view. The spatial coherence of rainfall and landslide data and the availability of long records are critical obstacles to successful rainfall threshold analysis.

Third is the lack of proper validation of rainfall ID curves. This is a key problem which has been identified to be a major limitation for the use of ID threshold models [e.g., *von Ruette et al.*, 2011; *Martelloni et al.*, 2012; *Segoni et al.*, 2013]. Performance in most studies is commonly evaluated only in calibration, by statistical performance measures for which no reference is provided and no extensive study of model robustness is carried out. *Gariano et al.* [2015], *Piciullo et al.* [2016], *Martelloni et al.* [2012], and *Segoni et al.* [2013] are examples of the few studies that conducted a calibration and validation of model performance, proposing a methodology useful for operational landslide warning systems.

In this work, we provide an analysis of the predictive performance of rainfall thresholds for shallow landslides in an alpine environment (Switzerland) in which we systematically address all of the above problems. Our main goals are (1) to formulate a robust statistical method to objectively define rainfall thresholds by considering different precipitation characteristics (maximum depth, peak intensity, duration, etc.) and their combination in the form of an ID threshold curve; (2) to investigate the additional information content of land surface erodibility and climate (mean daily precipitation) as predictors of landsliding, by comparing rainfall thresholds conditioned on surface erodibility and/or climate with independent thresholds for the entire country; and (3) to perform a validation of the rainfall threshold models by providing performance measures for reference cases and quantifying the sensitivity of estimated rainfall thresholds to individual wet-dry years and to the sample size in general.

The novelty of our work is that we have combined a new high-resolution gridded daily precipitation data set for Switzerland with a long landslide inventory with over 2000 recorded landslides with known timing over a 41 year period (1972–2012). This allowed us to perform analyses which were not possible previously, for example (a) we can use high-quality gridded precipitation and avoid allocating every landslide to the nearest raingauge, (b) we can include nontriggering events at landslide locations explicitly in the analysis and quantify false alarms, (c) we can define localized thresholds based on a large enough number of triggering events, and (d) we can develop reference cases in which we randomize the location and/or timing of landslides and perform meaningful regionalization experiments because our rainfall-landslide sample size is large enough. Although our rainfall thresholds pertain only to Switzerland, we think that the methods are

applicable to any location with sufficient data, and the results highlight the potential and limitations of using daily rainfall thresholds for landslide assessment in general.

2. Data and Methods

2.1. Rainfall and Landslide Data

The main data used in this study are the gridded mean daily precipitation *RhiresD* (Swiss Federal Office of Meteorology and Climatology MeteoSwiss) and the *Swiss flood and landslide damage database* (Swiss Federal Research Institute WSL). *RhiresD* consists of daily (6 A.M. to 6 A.M.) precipitation depths for approximately 2×2 km cells covering Switzerland since 1961. The gridded product is the result of the following steps explained in *Frei et al.* [2006]: the mean daily climatology is obtained following a modified version of PRISM [Schwarb, 2000], then the daily anomaly relative to the climatology is computed for each station and the corresponding anomaly field is interpolated in space (modified SYMAP interpolation algorithm [Shepard, 1984; Frei and Schär, 1998]). Finally, the interpolated relative anomalies are multiplied by the field of climatological means. Between 430 and 460 raingauges across the country are used in the interpolation. The *Swiss flood and landslide damage database* contains records of floods, debris flows, landslides, and rockfalls which caused damage to buildings, infrastructure or led to fatalities in Switzerland since 1972 [Hilker et al., 2009]. The records were taken from media reports, newspaper articles, administration records, and scientific monitoring. Only landslides with a known location and timing were included in our work ($n = 2271$). The overlapping period of the precipitation and landslide data sets on which the analysis was conducted spans 41 years from January 1972 to December 2012.

An example of a rainy day in August 2005 in Figure 1 shows heavy rainfall across central Switzerland accompanied by wide-spread landsliding. In a first step, we collocated the landslides to their corresponding precipitation cell in space (e.g., Figure 1b) for the entire study period 1972–2012. If multiple landslides occurred within one cell on the same day, these were considered all triggered by the same rainfall event. The matching resulted in 1562 unique cells in which landslides were recorded at least once, i.e., about 15% of the area of Switzerland. These are cells (areas) which we consider as susceptible to landsliding. The entire analysis was carried out only on the susceptible cells to avoid the inclusion of areas where landslides cannot happen for reasons unrelated to precipitation or would not be reported in the landslide database.

Figure 1. (a) Example of the precipitation depth on 22 August 2012 in mm/d (source: MeteoSwiss, *RhiresD*) and the locations of landslides recorded on that day and from the entire study period 1972–2012 (source: WSL, *Swiss flood and landslide damage database*). (b) Schematic of the collocation of landslides to precipitation cells in a neighborhood of nine cells. (c) Schematic of the assignment of precipitation events as triggering and nontriggering depending on landslide occurrence.

In order to investigate the impact of landslide database quality, we conducted the analysis on two selections of landslides: using all $n = 2271$ events and using a smaller selection which excluded landslides that have not been primarily triggered by rainfall. This selection of $n = 1670$ events removes those landslides where the triggering mechanism in the database is identified as something other than rainfall, e.g., snowmelt. We show mostly the results of the latter selection in this paper, but we also conduct comparisons of both selections to illustrate the impact of landslide database quality, i.e., the case where the landslide inventory is not purely rainfall-triggered but also includes events of unknown triggering.

2.2. Rainfall Threshold Model

To objectively define rainfall thresholds for the landslide susceptible cells,

Figure 2. Example of the division of precipitation events (consecutive days with precipitation above $P_{min} = 1$ mm/d) into triggering and nontriggering events and explanation of event parameters, considering the entire duration of the triggering event as well as only up to the day of the landslide (UTL).

we use an approach which consists of two steps: (1) the definition of key rainfall event properties which are important for generating landslides and the division of these events into landslide triggering and nontriggering and (2) the calibration of thresholds for rainfall event properties based on an objective maximization of predictive performance, i.e., correct prediction of triggering and nontriggering events. Details of these two steps in the rainfall threshold model are explained as follows.

The first step is to build the rainfall event history on the selected landslide susceptible cells and define event properties that may be relevant for landslide triggering (Figures 1c and 2). An independent event is a series of consecutive rainy days in which more than 1 mm/d of precipitation was measured. The minimum threshold of 1 mm/d was chosen so that annual totals would not be significantly reduced and realistic multiday event durations would be obtained. For each event, we compute the maximum daily intensity I_{max} (mm/d), total cumulative rainfall E (mm), duration D (day), and mean daily intensity $I_{mean} = E/D$ (mm/d). Next, events are divided into triggering and nontriggering subsets. An event is considered as triggering if at least one landslide was recorded in the corresponding cell during or on the day after rainfall, and as nontriggering otherwise. The day after is also considered because daily rainfall in the *RhiresD* data set is measured till 6 A.M. of the following day and therefore may contribute to soil wetting and landslide generation on that day. Finally, for the triggering events we consider two options for event properties (Figure 2): (a) taking the entire event duration regardless of the day of occurrence of the landslide (referred to as entire event) or (b) taking event properties only up to the day of the landslide (referred to as Up-To-Landslide or UTL). The reason for studying both options is to allow comparison with previous studies which mostly use option a, and at the same time to develop a methodology which can be used in a forecasting and warning mode based on the minimum required rain to generate landslides, which implies option b. Additionally, a comparison between the two options discerns the effect of considering entire events in the case when the timing of landslides is unknown (or inaccurate) in the landslide database.

The second step consists of selecting a threshold for rainfall that separates the observed triggering and non-triggering events with the best level of predictive performance. For this purpose we use the confusion matrix (contingency table) [Pearson, 1904], in which we count the true/false positives and true/false negatives from our rainfall threshold model and compute prevalence-independent goodness-of-fit measures which are commonly used in landslide assessment studies [e.g., Begueria, 2006; Staley et al., 2013; Gariano et al., 2015; Corsini and Mulas, 2016]. We find the rainfall threshold that maximizes the True Skill Statistic (TSS) as an integrative measure of landslide prediction performance. $TSS = s_e - (1 - s_p)$ is the difference between the true positive rate (sensitivity s_e) and the false alarm rate ($1 - \text{specificity}$ s_p) which are the two most important components for providing early warnings. TSS measures the performance of the rainfall threshold model beyond what can be achieved by a random guess when $TSS = 0$, with best possible performance when $TSS = 1$. The reason why it is important to use prevalence-independent statistical measures such as s_e , s_p in the TSS is that the sample

sizes of landslide triggering and nontriggering rainfall events are vastly different. Otherwise, the optimization of the rainfall threshold model would always be biased toward the more prevalent nontriggering events. TSS is a measure also known in the literature as the Peirce Skill Score [Peirce, 1884], Youden's index [Youden, 1950], or Hanssen and Kuipers discriminant [Hanssen and Kuipers, 1965].

A rainfall threshold is defined in two ways: (a) by maximizing TSS for every event property independently (I_{max} , E , D , I_{mean}) and (b) by maximizing TSS for a typical intensity-duration (ID) threshold curve for rainfall triggering in the form $I_{mean} = a \cdot D^{-b}$ [Caine, 1980]. The overall performance of a rainfall threshold model is also illustrated with the so-called receiver operating characteristic (ROC) curve [e.g., Fawcett, 2006] of s_e versus s_p for different rainfall thresholds. While the TSS identifies the threshold that has the highest predictive performance beyond a random choice, the area under the ROC curve (AUC) gives an overall impression of the predictive power of an event property, regardless of a specific threshold value. Step two of the rainfall threshold model analysis was performed for each precipitation property independently to find the appropriate threshold and associated TSS and AUC values, and also for the ID threshold curve to estimate the parameters a and b which maximized TSS.

2.3. Regionalization of Thresholds

Landslides are a complex geomorphic process caused by the interplay of triggering factors with the predisposition of slopes to failure due to their local soil properties and thickness, bedrock conditions, presence or absence of vegetation, slope steepness, etc. Due to the heterogeneity of predisposing factors, a single country-wide threshold for triggering rainfall might not be adequate. Therefore, we assessed the information content of two relevant properties—erodibility and climate—by comparing the single country-wide threshold with localized thresholds obtained in subregions with similar properties. The two chosen properties are first-order controls on the landscape: local erodibility of the soil was quantified from the geotechnical map [Kühni and Pfiffner, 2001], and the local climate was quantified by the long-term mean daily precipitation (MDP) at each landslide site and is indicative of local erosivity.

The erodibility map of Kühni and Pfiffner [2001], which captures different resistance against incision by glaciers and rivers and mass wasting by slope processes, was used to define four regions in Switzerland from very low to high erodibility. Long-term mean daily precipitation was computed from the RhiresD data set and Switzerland was also divided into four classes of intensity. As a rainfall event property, we chose I_{max} which was implemented in the rainfall threshold model to define 4×4 erodibility and MDP threshold combinations using I_{max} and normalized $I'_{max} = I_{max}/MDP$. For the latter, the I'_{max} threshold is determined for each region, and then each precipitation cell is assigned a unique I_{max} threshold depending on the local cell MDP ($I_{max} = I'_{max} \times MDP_{cell}$). We decided to consider I_{max} because using the ID power curve both slope and intercept can change. This would make the results less straightforward to interpret and also less robust as the number of triggering events (landslides) decreases with each regionalization. In summary, four different regionalization analyses were carried out and compared: 1 threshold for the entire country, 4 thresholds for the erodibility classes, 4 thresholds for the MDP classes, and 16 thresholds for the combinations of erodibility and MDP. To better understand the relationship between MDP, triggering rainfall and landslide probability, the Bayesian conditional probability of a landslide given a maximum rainfall intensity range was also estimated for each MDP region.

2.4. Methods for Testing and Validation

Testing and validation of the performance of the rainfall threshold model focused on three aspects: (1) we tested the performance of the model against cases in which the collocation of landslide-rainfall data was removed; (2) we tested the effect of nontriggering event prevalence and short sampling on the variability of calibrated rainfall thresholds; and (3) we conducted a traditional cross-validation experiment where thresholds were estimated by a leave-one-out procedure.

The first step in the testing of the performance of the rainfall threshold model was to develop reference cases where we disrupted the spatial and temporal coherence of rainfall and landslide occurrence. In the ROC space, we then tested the overall model performance (AUC) with the aim to judge if (and how much) is model performance augmented by spatial and temporal coherence. This test required resampling of the rainfall-landslide data on susceptible cells by randomizing the location of landslides and by randomizing the time (day) of their occurrence. In the first case, each observed landslide was assigned to a randomly

selected rainy cell on that day out of all susceptible cells. In the second case, each landslide was assigned to a random day in the study period in the same cell. In order to avoid assigning a landslide to a dry day, this selection was done on rainfall events in the cell. This assumes that landslides occur always on the final day of the triggering event. Each resampling experiment was repeated 100 times and the performance statistics, ROC curves and their spread were reported.

The second step in the testing of the performance of the model was to quantify the prevalence problem, i.e., the uncertainty in the thresholds (and ROC curves) that is due to unequal triggering and nontriggering rainfall event sample sizes. To this end we compared the results with all observed rainfall events (for $n = 1670$ triggering events and over 3 million nontriggering events) with a case in which we sampled also exactly $n = 1670$ nontriggering events at random. The duration of the record was also examined by sampling N randomly selected years ($N = 5$ or $N = 20$ years) from the total sample of 41 years in the period 1972–2012. Like in the previous case, each resampling experiment was repeated 100 times and the performance statistics and ROC curves and their spread were reported.

The third step is the validation of the rainfall threshold model. This is important for judging how good predictions with the model are when used for instance in a forecasting mode. We performed validation by a jack-knife resampling and by cross-validation. For jack-knife resampling we removed each year in turn from the data set and carried out the calibration with the remaining 40 years, as well as with only the removed year. For cross-validation, independent samples of N consecutive years were used for calibration and the validation was performed on the remaining $41 - N$ years. In all cases the thresholds and their uncertainty were reported. The critical assumption behind the validation performed here is that the rainfall-landslide records are both stationary and ergodic.

3. Results

3.1. Rainfall-Landslide Thresholds for Switzerland

The calibrated thresholds both for individual event parameters and for the ID curve are shown in Table 1. Concerning the independent event properties I_{max} , E , I_{mean} , and D , it can be concluded that maximum daily event intensity provides the best predictive performance (TSS = 0.64, AUC = 0.9 for UTL), followed closely by total event depth. Event duration has by far the lowest predictive performance by itself. The country-wide threshold for maximum daily intensity was $I_{max} = 28$ mm/d, and with this threshold 76% of all landslides were predicted correctly and the false alarm rate was 12% considering event properties only up to the time of the landslide. The performance was consistently better for all event properties when the entire event duration was considered.

Table 1. Calibrated Thresholds for Individual Rainfall Event Variables and the Corresponding Performance Statistics Obtained Considering Each Variable Separately and in the Form of an Intensity-Duration (ID) Curve^a

	Variable	Units	Th	Spec	Sens	TSS	AUC
UTL	Maximum intensity	mm/d	28.3	0.88	0.76	0.64	0.90
	Total rainfall	mm	43.0	0.83	0.76	0.59	0.87
	Duration	d	3	0.65	0.58	0.23	0.65
	Mean intensity	mm/d	15.4	0.86	0.80	0.65	0.90
	Max intensity/MDP	mm/d	6.8	0.85	0.80	0.66	0.90
	ID curve		$a = 18.3; b = 0.21$	0.85	0.82	0.67	
Entire events	ID curve (I_{mean}/MDP)		$a = 5; b = 0.22$	0.85	0.83	0.69	
	Maximum intensity	mm/d	29.1	0.89	0.80	0.68	0.92
	Total rainfall	mm	47.4	0.86	0.82	0.68	0.91
	Duration	d	4	0.79	0.65	0.45	0.79
	Mean intensity	mm/d	13.7	0.81	0.80	0.62	0.89
	Max intensity/MDP	mm/d	7.2	0.87	0.83	0.70	0.92
UTL, all landslides	ID curve		$a = 29.7; b = 0.65$	0.85	0.85	0.69	
	Maximum intensity	mm/d	25.0	0.85	0.73	0.57	0.86
	Total rainfall	mm	43.0	0.83	0.71	0.54	0.85
	Duration	d	3	0.65	0.59	0.24	0.65
	Mean intensity	mm/d	14.6	0.84	0.74	0.58	0.86
	Max intensity/MDP	mm/d	6.7	0.85	0.75	0.59	0.87
	ID curve		$a = 19.3; b = 0.25$	0.85	0.75	0.60	

^aThe results are reported for complete rainfall triggered events (entire events) and for up to the day of the landslide (UTL events). Also listed are UTL results for all landslides, including some landslides caused by triggers other than rainfall.

Figure 3. Intensity-duration plots with threshold lines from Guzzetti *et al.* [2007], Berti *et al.* [2012], and Zimmermann [1997], and those computed in this paper from data considering entire events (best TSS) or the triggering events only up to the day of the landslide (best TSS, UTL). The color maps show the relative frequency of the triggering events (a) entire events, (b) UTL, and (c) nontriggering events. Individual landslide triggering events are also shown as dots in Figures 3a and 3b.

As expected, the ID curves produce a better overall performance than individual event properties (Table 1). This is because the ID curve includes both event intensity and duration in defining the threshold and this combination allows to account for landslides triggered by shorter (daily) intense events as well as long lasting multiday events with lower intensity. The calibrated ID curve manages to separate well the hotspots of relative frequency of triggering and nontriggering events in the duration-mean intensity space (Figure 3). Three additional ID curves from the literature were chosen for comparison: Guzzetti *et al.* [2007] using a Bayesian inference method and considering triggering events only for the climatic region of Switzerland, Berti *et al.* [2012] as a lower envelope curve defined at the daily scale and Zimmermann [1997] for debris flows in Switzerland. The latter is obtained considering only extreme events, both triggering and nontriggering, and therefore is much higher than the others. The fitted ID curve to the entire events in this study is very similar to the curve of Guzzetti *et al.* [2007].

Comparing the ID curve plot for entire events (Figure 3a) with that which considers rainfall only up to the day of the landslide (UTL) (Figure 3b) shows that many landslides were triggered on the first day of longer storms. This strongly affects the ID curve parameters even though the overall performance of the model remains the same (Table 1). It is also likely that most of these 1 day events are summer convective storms which have a duration much less than 1 day [e.g., Molnar and Burlando, 2008], and therefore cannot be captured with daily data, or that antecedent precipitation played a significant role in their triggering. This is clearly a limitation of the daily time resolution used in our analysis.

3.2. The Impact of Regionalization

The single rainfall threshold for the entire country was compared with thresholds estimated for four classes (regions) of erodibility and four classes (regions) of mean daily precipitation (MDP) for I_{max} (Figure 4). A better separation of the triggering and nontriggering distributions of I_{max} is achieved in regions of low/very low erodibility (boxplots of triggering and nontriggering events in Figure 4) and the calibrated thresholds are higher there. In contrast in erodible soils less rainfall is needed to generate surface erosion and trigger landsliding. In fact, most landslides occur in the high erodibility region of Switzerland. A better separation of the triggering and nontriggering distributions of I_{max} is also achieved in wetter regions and rainfall thresholds clearly increase with MDP. The suggestion that higher rainfall is required to trigger landslides in

Figure 4. Boxplots of triggering (red) and nontriggering (green) maximum daily intensity for the entire country (allCH) and each region defined based on mean daily precipitation (mdp: low, medium low, medium high, high) or on erodibility (erodibility: very low, low, medium, high). The threshold I_{max} obtained for each class is represented with a red horizontal line. The boxplot graphs do not include outliers.

threshold computed over the entire country (Figure 6). The highest discretization in space into the 4×4 erodibility and MDP classes resulted in the highest TSS, mostly due to increasing sensitivity of the predictions (i.e., the number of landslides correctly predicted), while specificity remained roughly the same. Normalizing rainfall by MDP also slightly improved performance. However, the increase in performance by regionalization is small compared to the loss of performance that occurs when all events, even those where the triggering is not only rainfall, are included in the analysis (Figure 6, right). Reducing the quality of the landslide data set leads to a drop in TSS from 0.64 to 0.57 for I_{max} and from 0.66 to 0.59 for I_{max}/MDP . This effect is much greater than any gain by regionalization and points to the importance of obtaining good landslide inventories.

The estimated I_{max} thresholds and their exceedance probability in the period 1972–2012 for the case of a single threshold for the entire country and 16 regions given by combinations of erodibility and mean daily precipitation were compared (Figure 7). For the latter, 16 I_{max}/MDP thresholds were defined for the different

Figure 5. Posterior probability of a landslide given the maximum daily intensity I_{max} [mm/d] range for each of the four MDP [mm/d] classes.

wet regions was tested further by computing the posterior probability of landsliding conditioned on I_{max} for the four different MDP classes. A precipitation event with the same I_{max} has a significantly higher probability of generating a landslide in a region with lower MDP, at least up to $I_{max} = 80$ mm/d (Figure 5). The balance between rates of soil formation and erosion in landscapes shaped by landsliding suggests that the availability of sufficient soil cover may limit the frequency of landsliding in highly erosive regions (high MDP). This balance is implicitly assumed when considering rainfall event variables normalized by MDP [e.g., Cannon, 1988; Guzzetti *et al.*, 2008; Guidicini and Iwasa, 1977], as well as when thresholds are based solely on precipitation distributions [e.g., Pedrozzi, 2004].

The rainfall threshold model performance for I_{max} and I_{max}/MDP was slightly improved by regionalization compared to the single

regions separately and then the local MDP of each cell was used to obtain the corresponding I_{max} threshold, so in fact there is a unique threshold for every cell rather than region. The thresholds are indeed lower in the northern pre-Alpine regions and mountain valleys, while they are highest in the central Alps and southern Switzerland (Tessin) where precipitation tends to be highest. We propose that this level of spatial resolution is adequate for regional-scale mapping of rainfall thresholds for first-order landslide early warning systems based on forecasts of daily rainfall.

3.3. Rainfall Threshold Performance Assessment

Most previous studies ignore the validation of rainfall threshold models for landslide triggering mainly because

Figure 6. True skill statistic, specificity, and sensitivity values associated with the different studied cases: all CH (considering the entire country), four erod. (four regions defined by erodibility), four mdp (four regions defined by mean daily precipitation), 16 erod. and mdp (16 classes given by the combinations of each erodibility and MDP class), all CH all land (analysis on the entire country, but including also landslides associated with snowmelt or other triggers).

Interestingly, the ROC curves with randomized locations still show some residual predictive power, compared to the case where the timing of landslides was randomized. As expected the latter followed the random model for which $s_e = 1 - s_p$. The residual predictive power comes from the fact that many landslides in our data set occur during long-lasting storms with large spatial extent over Switzerland. In this case, randomizing the location of some landslides results in placing them in cells that also have high (triggering) rainfall.

The results also show that sample size (record length) is indeed influencing performance statistics and leads to a rather large variability around the ROC curve (compare the spread of 5 and 20 year sampling) even though there is no evident bias. Interestingly, the prevalence problem is much smaller than the sampling duration problem, meaning that when the same number of nontriggering and triggering events are sampled, the spread in ROC curves is rather small (similar to sampling all events for 20 years). For example, the mean threshold for the prevalence-corrected ROC curves was 28.3 mm/d (with a standard deviation of 2.1 mm/d), which is identical to the threshold obtained considering all nontriggering events (Table 1). This gives credibility to the TSS as a robust performance statistic for landslide analysis.

Second, we conducted a validation where thresholds were estimated using only events from some years and prediction errors were computed for the remaining years (Figure 9). The number of landslides in Switzerland does not show a rising tendency in time (Figure 9a), only the year 2005 is an outlier in which almost 200 landslides occurred, many of them in one August event (Figure 1). The boxplot of I_{max} in nontriggering rainfall events also shows that there is no tendency in time, with a median $I_{max} \sim 10$ mm/d throughout the entire period 1972–2012 and low interannual variability (Figure 9d). However, strong interannual variability is evident in I_{max} in triggering rainfall events (Figure 9b) and consequently in the thresholds defined for each year separately (Figure 9c). In some years the I_{max} in landslide triggering events was almost 3 times greater than the value estimated for the entire period.

The jack-knife validation is summarized in terms of the I_{max} thresholds estimated from individual years and from the sample with each year removed (the line in Figure 9c gives the threshold for all years). The results show that the thresholds are reasonably robust to interannual variability, even removing the most extreme year 2005 led to a change in the calibrated threshold of only 10%. Similar conclusions can be drawn from the cross-validation analysis (Figure 10). The best I_{max} threshold is obtained in calibration for different independently sampled number of years (Figures 10a), validation is carried out on the remaining years, and the resulting statistics are reported for both sets (Figures 10b–10d). It is evident that the ROC and TSS method

the landslide sample size is usually small. In this study, we have the advantage of a large sample size which allows us to perform a proper validation experiment.

First, we show the effect of the disruption of colocation of landslide-rainfall data, short sampling, and prevalence on the ROC curves (Figure 8). Each of the cases explained in section 2.4 is resampled $n = 100$ times resulting into 100 ROC curves for I_{max} for each test. The results, compared to those from the full data set (thick black line in Figure 8), show that randomizing the location or the timing with respect to the actual landslide database leads to a performance that is far worse than any of models with colocated data even if only 5 years were used. The maximum TSS is 0.40 (maximum AUC 0.76) for random locations whereas for the random 5 years the minimum TSS is 0.56 (minimum AUC 0.85). Interest-

Figure 7. Thresholds for I_{max} (top row) and its exceedance probability (bottom row) for one threshold over the entire country (left column) or 16 regions based on mean daily precipitation and erodibility (right column) using UTL events.

used for identifying thresholds is robust: the median TSS during validation is always above 0.6 (0.64 in calibration) and the performance both in terms of sensitivity (correct prediction of landslides) as well as specificity (correct prediction of no landslides) does not drop with small sample size.

Figure 8. ROC curve of maximum daily intensity in the analysis UTL over the entire country (all events) and 100 ROC curves for each sampling: selecting as many nontriggering events as triggering events (same # trig and nontrig), sampling 20 or 5 random years, randomizing the landslide location (random location) and their day of occurrence (random time).

4. Discussion

Regionalization of rainfall thresholds for landsliding based on geology, lithology, mean rainfall, land use, or zones with homogeneous hydrogeomorphological properties has been a target in previous studies [Tiranti and Rabuffetti, 2010; Peruccacci et al., 2012; Rosi et al., 2012; Vennari et al., 2014; Gariano et al., 2015]. In most of these works no conclusions on the effect of localizing the threshold could be drawn because of the limited amount of landslides in individual subregions. Some of the studies in fact showed that the added value of regionalizing thresholds was minimal. Also in our case, despite the number of landslides was largely sufficient (for MDP minimum 361 landslides per region, for erodibility minimum 184 landslides per region), the effect of regionalization on the rainfall threshold model performance was small (Figure 6). Nevertheless, clear and meaningful signals were present in the

Figure 9. Results of the validation on individual years as well as jack-knifing with one year removed from calibration: (a) the number of landslides per year, (b) the boxplot of triggering event I_{max} (maximum daily intensity) for each year, (c) the best I_{max} threshold for each year and for the remaining 40 years, and (d) the boxplot of the nontriggering event I_{max} for each year. The boxplot graphs do not include outliers.

regions, whereas that of triggering events varies (Figure 4). Therefore, in regions where the rainfall required for landslide initiation is high (wet climate or low erodibility) the separation between triggering and nontriggering becomes easier while the opposite is true in dry and highly erodible regions, compared to having a single threshold for the entire country.

The TSS is only one of many possible choices for model performance as an optimization criterion. It has several advantages, e.g., it accounts for both triggering and nontriggering events, it is prevalence independent, and contains hits and false alarms which are both measures relevant for early warning systems. However, because rainfall triggers landslides only occasionally, a 1% increase in the hit rate corresponds to about 17 more landslides predicted correctly, while a 1% increase in the false alarm rate corresponds to about 36,000 false alarms country wide. This discrepancy is also the reason why removing a few low-intensity triggering events which possibly could be miss-classifications from the data set will result in increasing sensitivity and TSS, but will likely not lead to a large change in the rainfall threshold itself. In applications for early warning systems in particular it may be better to consider giving different weights to specificity and sensitivity in the sense of setting different costs and penalties for false and missed alarms.

optimal rainfall thresholds themselves and their variability between regions, despite marginal increases in TSS when moving from a single threshold for the entire country to 16 regions. The spatial trends observed in the I_{max} threshold values and their consistency with expectations regarding the effects of erodibility and climate are indicating that there is added value to regionalization.

Possible explanations for the small effect of regionalization on TSS are several. First, the range of possible TSS values is much smaller than the theoretical range 0–1. In fact, even the experiment with randomized locations of landslides resulted in TSS equal to about 0.4. Also, the highest value that TSS can reach is certainly less than 1 due to the inherent stochasticity in landslide generation and the inadequacy of the simple rainfall threshold model. Misclassification may affect strong nontriggering events which in fact led to landslides which were not reported, or low-intensity triggering events caused by factors other than rainfall or by antecedent wetness conditions that were ignored. Second, the distribution of nontriggering events is relatively similar among different

Figure 10. Results of the cross-validation: (a) best threshold as a function of number of years used for the calibration, (b) specificity, (c) sensitivity, and (d) TSS associated with each validation that uses the threshold in Figure 10a to make predictions in the remaining years not used in calibration.

olds as suggested in previous studies [Destro *et al.*, 2017; Marra *et al.*, 2016; Nikolopoulos *et al.*, 2014, 2015]. Although it aims to reduce systematic biases due to under-representativeness of high altitudes, still a bias of about 10–15% in extremes is expected based on cross-validation tests carried out considering the density of the Swiss raingauge network [Isotta *et al.*, 2014]. Taking the closest raingauge to a landslide location, as typically done, is also possible, and we found that about 50% of the landslides occurred within 4 km of a nearest raingauge (80% within 7 km). We repeated the analysis with the rainfall data at the nearest raingauge and found that indeed the performance worsened only marginally (ROC). However, for predictive purposes, it is advantageous to work with gridded data to have a coherent spatial expression of rainfall triggering in a region. In future work, we plan to apply the same methodology with a highly resolved (5 min, 1 × 1 km) rainfall data set that combines raingauge measurements with radar observations and should therefore be able to capture also the local thunderstorms which a raingauge network and RhiresD might miss.

5. Conclusions

A gridded daily precipitation data set for Switzerland (RhiresD, MeteoSwiss) was combined with a landslide inventory containing over 2000 recorded landslides (Swiss flood and landslide damage database, WSL) over a 41 year period 1972–2012 to analyze rainfall events which lead to landsliding. The spatially resolved and continuous precipitation data and the large landslide sample size allowed us to colocate triggering rainfall to the landslides, to develop distributions of triggering and nontriggering rainfall event properties at susceptible sites, and to exploit them in determining rainfall thresholds for landslide triggering and validate their performance. Our main conclusions are:

1. The ROC method and True Skill Statistic TSS, which balances hits and false alarms, were successful in defining rainfall thresholds for landsliding. The best predictive performance was obtained by the intensity-duration threshold curve, followed by peak daily intensity I_{max} and mean event intensity I_{mean} as event properties. Event duration by itself has very low predictive power.
2. The regionalization of I_{max} rainfall thresholds based on surface erodibility and local climate determined by mean daily precipitation at landslide-susceptible sites showed that local I_{max} thresholds vary meaningfully in space compared to a single country-wide threshold. Wetter local climate and lower erodibility led to higher rainfall thresholds. However, the improvement in TSS performance due to regionalization was small and much lower than what can be achieved by having a high-quality landslide database.
3. Reference cases in which the landslide locations and timing were randomized and the landslide sample size was reduced showed the sensitivity of the I_{max} rainfall threshold model. A residual predictive power was present even when landslide locations were randomized because landslide-generating storms

Two critical points of this analysis concern the precipitation data set. The first aspect is the temporal daily resolution. In fact, sub-daily precipitation data are more suitable to capture the short convective events that can initiate a landslide [e.g., Schilirò *et al.*, 2015] and are therefore likely to have a higher predictive power. On the other hand, using a higher time resolution typically reduces significantly the number of landslide events available for the analysis because accurate occurrence timing is required. In addition, the results presented here showed that the model still has predictive power also when used at a daily scale. Finally, the fact that daily precipitation forecasts are more reliable is relevant in an early warning context. The second critical aspect concerns the spatial accuracy. Because RhiresD is an interpolated product, it could lead to underestimation of the extreme events and therefore the thresh-

generally have a large spatial extent. Short records led to larger variability in ROC curves and TSS values but no bias. Jack-knife and cross-validation experiments demonstrated that the rainfall threshold model is robust.

Our results reported here highlight the potential of using rainfall I-D threshold curves and I_{max} threshold values for predicting the occurrence of landslides on a country or regional scale. Two main developments are foreseen for future work in Switzerland: extension to a subdaily timescale which would allow to capture the triggering caused by short duration intense convective events and inclusion of antecedent soil wetness status in the form of a hydrological soil moisture accounting scheme/model. Taking into account antecedent soil wetness will also remove the need to define precipitation events and therefore to subjectively choose interarrival time and minimum rainfall required to identify storms.

Acknowledgments

This research was funded by the Swiss National Science Foundation grant 165979 awarded to P. M. The rainfall data were provided by the Swiss Federal Office of Meteorology and Climatology MeteoSwiss (available for research purposes upon request). The Swiss Federal Research Institute WSL provided the landslide data (available for research purposes upon request). E. L. conducted the analysis. P. M., E. L., and B. W. M. conceived the research. All authors contributed to writing the paper.

References

- Beguería, S. (2006), Validation and evaluation of predictive models in hazard assessment and risk management, *Nat. Hazards*, 37(3), 315–329.
- Berti, M., M. Martina, S. Franceschini, S. Pignone, A. Simoni, and M. Pizzolo (2012), Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach, *J. Geophys. Res.*, 117, F04006, doi:10.1029/2012JF002367.
- Brunetti, M. T., S. Peruccacci, M. Rossi, S. Luciani, D. Valigi, and F. Guzzetti (2010), Rainfall thresholds for the possible occurrence of landslides in Italy, *Nat. Hazards Earth Syst. Sci.*, 10(3), 447–458, doi:10.5194/nhess-10-447-2010.
- Brunetti, M. T., et al. (2013), Rainfall thresholds for possible occurrence of shallow landslides and debris flows in Italy, in *Dating Torrential Processes on Fans and Cones: Methods and Their Application for Hazard and Risk Assessment*, pp. 327–339, Springer Science and Business Media, Dordrecht, doi:10.1007/978-94-007-4336-6_22.
- Caine, N. (1980), The rainfall intensity: Duration control of shallow landslides and debris flows, *Geogr. Ann. Ser. A*, 62(1/2), 23–27.
- Cannon, S. (1988), Regional rainfall-threshold conditions for abundant debris-flow activity, *U.S. Geol. Surv. Prof. Pap.*, 1434, 35–42.
- Corsini, A., and M. Mulas (2016), Use of roc curves for early warning of landslide displacement rates in response to precipitation (Piagneto landslide, Northern Apennines, Italy), *Landslides*, 14, 1241–1252, doi:10.1007/s10346-016-0781-8.
- Crozier, M. J. (1999), Prediction of rainfall-triggered landslides: A test of the antecedent water status model, *Earth Surf. Processes Landforms*, 24(9), 825–833.
- Destro, E., F. Marra, E. I. Nikolopoulos, D. Zoccatelli, J. D. Creutin, and M. Borga (2017), Spatial estimation of debris flows-triggering rainfall and its dependence on rainfall return period, *Geomorphology*, 278, 269–279, doi:10.1016/j.geomorph.2016.11.019.
- Fawcett, T. (2006), An introduction to ROC analysis, *Pattern Recognit. Lett.*, 27(8), 861–874, doi:10.1016/j.patrec.2005.10.010.
- Frei, C., and C. Schär (1998), A precipitation climatology of the Alps from high-resolution rain-gauge observations, *Int. J. Climatol.*, 18(8), 873–900, doi:10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9.
- Frei, C., R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale (2006), Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, *J. Geophys. Res.*, 111, D06105, doi:10.1029/2005JD005965.
- Gariano, S. L., M. T. Brunetti, G. Iovine, M. Melillo, S. Peruccacci, O. Terranova, C. Vennari, and F. Guzzetti (2015), Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy, *Geomorphology*, 228, 653–665, doi:10.1016/j.geomorph.2014.10.019.
- Glade, T. (2000), Modelling landslide-triggering rainfalls in different regions of New Zealand-The Soil Water Status Model, *Z. Geomorph. N.F.*, 122, 63–84.
- Glade, T., M. Crozier, and P. Smith (2000), Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical “antecedent daily rainfall model,” *Pure Appl. Geophys.*, 157(6–8), 1059–1079.
- Guidicini, G., and O. Iwasa (1977), Tentative correlation between rainfall and landslides in a humid tropical environment, *Bull. Int. Assoc. Eng. Geol.*, 16(1), 13–20.
- Guzzetti, F., S. Peruccacci, M. Rossi, and C. P. Stark (2007), Rainfall thresholds for the initiation of landslides in central and southern Europe, *Meteorol. Atmos. Phys.*, 98(3–4), 239–267, doi:10.1007/s00703-007-0262-7.
- Guzzetti, F., S. Peruccacci, M. Rossi, and C. P. Stark (2008), The rainfall intensity–duration control of shallow landslides and debris flows: An update, *Landslides*, 5(1), 3–17.
- Hanssen, A., and W. Kuipers (1965), *On the Relationship Between the Frequency of Rain and Various Meteorological Parameters: With Reference to the Problem of Objective Forecasting*, Staatsdrukkerij-en Uitgeverijbedrijf, The Hague.
- Hilker, N., A. Badoux, and C. Hegg (2009), The Swiss flood and landslide damage database 1972–2007, *Nat. Hazards Earth Syst. Sci.*, 9(3), 913–925.
- Isotta, F. A., et al. (2014), The climate of daily precipitation in the Alps: Development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data, *Int. J. Climatol.*, 34(5), 1657–1675, doi:10.1002/joc.3794.
- Kjekstad, O., and L. Highland (2009), Economic and social impacts of landslides, in *Landslides—Disaster Risk Reduction*, pp. 573–587, Springer, Verlag Berlin Heidelberg.
- Klose, M. (2015), *Landslide Databases as Tools for Integrated Assessment of Landslide Risk*, Springer, International Publishing Switzerland, doi:10.1007/978-3-319-20403-1.
- Kühni, A., and O. Pfiffner (2001), The relief of the Swiss Alps and adjacent areas and its relation to lithology and structure: Topographic analysis from a 250-m DEM, *Geomorphology*, 41(4), 285–307.
- Marra, F., E. I. Nikolopoulos, J. D. Creutin, and M. Borga (2016), Space-time organization of debris flows-triggering rainfall and its effect on the identification of the rainfall threshold relationship, *J. Hydrol.*, 541, 246–255, doi:10.1016/j.jhydrol.2015.10.010.
- Martelloni, G., S. Segoni, R. Fanti, and F. Catani (2012), Rainfall thresholds for the forecasting of landslide occurrence at regional scale, *Landslides*, 9(4), 485–495.
- Meyer, N. K., A. V. Dyrddal, R. Frauenfelder, B. Eitzelmüller, and F. Nadim (2012), Hydrometeorological threshold conditions for debris flow initiation in Norway, *Nat. Hazards Earth Syst. Sci.*, 12(10), 3059–3073, doi:10.5194/nhess-12-3059-2012.
- Molnar, P., and P. Burlando (2008), Variability in the scale properties of high-resolution precipitation data in the Alpine climate of Switzerland, *Water Resour. Res.*, 44, W10404, doi:10.1029/2007WR006142.

- Nikolopoulos, E. I., S. Crema, L. Marchi, F. Marra, F. Guzzetti, and M. Borga (2014), Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence, *Geomorphology*, 221, 286–297, doi:10.1016/j.geomorph.2014.06.015.
- Nikolopoulos, E. I., M. Borga, J. D. Creutin, and F. Marra (2015), Estimation of debris flow triggering rainfall: Influence of rain gauge density and interpolation methods, *Geomorphology*, 243, 40–50, doi:10.1016/j.geomorph.2015.04.028.
- Pearson, K. (1904), *Mathematical Contributions to the Theory of Evolution. XIII. On the Theory of Contingency and Its Relation to Association and Normal Correlation*, *Drapers' Co. Res. Mem. Biometric Ser.*, vol. 1, Drapers' Co, London.
- Pedrozzì, G. (2004), Triggering of landslides in canton ticino (switzerland) and prediction by the rainfall intensity and duration method, *Bull. Eng. Geol. Environ.*, 63(4), 281–291, doi:10.1007/s10064-004-0240-y.
- Peirce, C. S. (1884), The numerical measure of the success of predictions, *Science*, 4(93), 453–454.
- Peruccacci, S., M. T. Brunetti, S. Luciani, C. Vennari, and F. Guzzetti (2012), Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central Italy, *Geomorphology*, 139, 79–90.
- Picciullo, L., S. L. Gariano, M. Melillo, M. T. Brunetti, S. Peruccacci, F. Guzzetti, and M. Calvello (2016), Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides, *Landslides*, 2016, 1–14, doi:10.1007/s10346-016-0750-2.
- Ponziani, F., C. Pandolfo, M. Stelluti, N. Berni, L. Brocca, and T. Moramarco (2012), Assessment of rainfall thresholds and soil moisture modeling for operational hydrogeological risk prevention in the Umbria region (central Italy), *Landslides*, 9(2), 229–237, doi:10.1007/s10346-011-0287-3.
- Rosi, A., S. Segoni, F. Catani, and N. Casagli (2012), Statistical and environmental analyses for the definition of a regional rainfall threshold system for landslide triggering in Tuscany (Italy), *J. Geogr. Sci.*, 22(4), 617–629, doi:10.1007/s11442-012-0951-0.
- Salvati, P., C. Bianchi, M. Rossi, and F. Guzzetti (2010), Societal landslide and flood risk in Italy, *Nat. Hazards Earth Syst. Sci.*, 10, 465–483, doi:10.5194/nhess-10-465-2010.
- Schilirò, L., C. Esposito, and G. Scarascia Mugnozza (2015), Evaluation of shallow landslide-triggering scenarios through a physically based approach: An example of application in the southern Messina area (northeastern Sicily, Italy), *Nat. Hazards Earth Syst. Sci.*, 15(9), 2091–2109.
- Schmidt, J., G. Turek, M. P. Clark, M. Uddstrom, and J. R. Dymond (2008), Probabilistic forecasting of shallow, rainfall-triggered landslides using real-time numerical weather predictions, *Nat. Hazards Earth Syst. Sci.*, 8(2), 349–357, doi:10.5194/nhess-8-349-2008.
- Schuster, R. L. (1996), Socioeconomic significance of landslides, *Landslides: Investigation and Mitigation*, *Transp. Res. Board Spec. Rep.* 247, pp. 12–35, Natl. Acad. Press, Washington, D. C.
- Schwarb, M. (2000), The Alpine precipitation climate, PhD dissertation, p. 131, Swiss Fed. Inst. of Technol. Zurich, Zurich, Switzerland.
- Segoni, S., L. Leoni, A. Benedetti, F. Catani, G. Righini, G. Falorni, S. Gabellani, R. Rudari, F. Silvestro, and N. Rebora (2009), Towards a definition of a real-time forecasting network for rainfall induced shallow landslides, *Nat. Hazards Earth Syst. Sci.*, 9(6), 2119–2133.
- Segoni, S., G. Martelloni, D. Lagomarsino, R. Fanti, and F. Catani (2013), An operational warning system for the forecasting of landslide occurrence at regional scale, in *Landslide Science and Practice*, pp. 611–617, Springer, Verlag Berlin Heidelberg.
- Shepard, D. S. (1984), Computer mapping: The SYMAP interpolation algorithm, in *Spatial Statistics and Models*, pp. 133–145, Springer Netherlands, Dordrecht.
- Stahli, M., et al. (2015), Monitoring and prediction in early warning systems for rapid mass movements, *Nat. Hazards Earth Syst. Sci.*, 15(4), 905–917, doi:10.5194/nhess-15-905-2015.
- Staley, D. M., J. W. Kean, S. H. Cannon, K. M. Schmidt, and J. L. Laber (2013), Objective definition of rainfall intensity-duration thresholds for the initiation of post-fire debris flows in southern California, *Landslides*, 10(5), 547–562, doi:10.1007/s10346-012-0341-9.
- Stevenson, P. (1977), An empirical method for the evaluation of relative landslide risk, *Bull. Int. Assoc. Eng. Geol.*, 16(1), 69–72.
- Tiranti, D., and D. Rabuffetti (2010), Estimation of rainfall thresholds triggering shallow landslides for an operational warning system implementation, *Landslides*, 7(4), 471–481, doi:10.1007/s10346-010-0198-8.
- Trezzini, F., G. Giannella, and T. Guida (2013), *Landslide and Flood: Economic and Social Impacts in Italy*, vol. 2, pp. 171–176, Springer, Verlag Berlin Heidelberg.
- Vennari, C., S. L. Gariano, L. Antronico, M. T. Brunetti, G. Iovine, S. Peruccacci, O. Terranova, and F. Guzzetti (2014), Rainfall thresholds for shallow landslide occurrence in Calabria, southern Italy, *Nat. Hazards Earth Syst. Sci.*, 14(2), 317–330, doi:10.5194/nhess-14-317-2014.
- von Ruetten, J., A. Papritz, P. Lehmann, C. Rickli, and D. Or (2011), Spatial statistical modeling of shallow landslides—validating predictions for different landslide inventories and rainfall events, *Geomorphology*, 133(1), 11–22.
- Youden, W. J. (1950), Index for rating diagnostic tests, *Cancer*, 3(1), 32–35, doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.
- Zimmermann, M. (1997), *Murganggefahr und Klimaänderung-ein GIS-basierter Ansatz*, VDF Hochschulverlag AG, Zürich, Switzerland.