

Cost-sensitive rainfall thresholds for shallow landslides

Abstract The risk management of rainfall-induced landslides requires reliable rainfall thresholds to issue early warning alerts. The practical application of these thresholds often leads to misclassifications, either false negative or false positive, which induce costs for the society. Since missed-alarm (false negative) and false-alarm (false positive) cost may be significantly different, it is necessary to find an optimal threshold that accounts for and minimises such costs, tuning the false-alarm and missed-alarm rates. In this paper, we propose a new methodology to develop cost-sensitive rainfall thresholds, and we also analyse several factors that produce uncertainty, such as the accuracy of rainfall intensity values at landslide location, the time of occurrence, the minimum rainfall amount to define the non-triggering event, and the variability of cost scenarios. Starting from a detailed mapping of landslides that occurred during five large-scale rainfall events in the Italian Central Alps, we first developed rainfall threshold curves with a ROC-based approach by using both rain gauge and bias-adjusted weather radar data. Then, based on a reference cost scenario in which we quantified several cost items for both missed alarms and false alarms, we developed cost-sensitive rainfall threshold curves by using cost-curve approach (Drummond and Holte 2000). Finally, we studied the sensitivity of cost items. The study confirms how important is the information regarding rainfall intensity at the landslide site for the development of rainfall thresholds. Although the use of bias-corrected radar strongly improves these values, a large uncertainty related to the exact time of landslide occurrence still remains, negatively affecting the analysis. Accounting for the different missed-alarm and false-alarm misclassification costs is important because different combinations of these costs make an increase or decrease of the rainfall thresholds convenient. In our reference cost scenario, the most convenient threshold is lower than ROC-based thresholds because it seeks to minimise the number of missed alarms, whereas the missed-alarm costs are almost seven times greater than false-alarm costs. However, for different cost scenarios, threshold may vary significantly, as much as half an order of magnitude.

Keywords Shallow landslides · Rainfall · ROC · Cost curve · Thresholds

Introduction

Rainfall is one of the most significant triggering factors for shallow landslides. Although many physically based, empirical and probabilistic approaches have been proposed in the literature (Campbell 1974; Pierson 1980; Larsen and Simon 1993; Montgomery and Dietrich 1994; Crozier and Glade 1999; Wiczorek et al. 2000), the prediction of landslides triggered by rainfall is still problematic to the complexity and the variability (in space and time) and scale dependency of rainfall and other controlling factors such as soil depth, soil resistance parameters, and soil hydraulic parameters. However, while the prediction of the exact location and timing of failure is almost impossible, a

prediction of critical rainfall conditions that may lead to landsliding is reasonably possible, especially for large rainfall events that trigger tens or hundreds of landslides. This case of large rainfall events may lead to critical emergency conditions that are extremely relevant for administrators that are called to issue alerts and manage civil protection actions.

A necessary condition for risk management through alerting is the availability of a rainfall threshold, that defines a level of rainfall needed for landslide triggering, and that may be used for issuing an alert. The most popular approach for the definition of a threshold is to define a rainfall intensity-duration (ID) threshold curve which accounts for both the intensity and duration of events that can trigger landslides (Stevenson 1977; Caine 1980; Glade et al. 2000; Wiczorek et al. 2000; Crosta and Frattini 2001; Guzzetti et al. 2007; Frattini et al. 2009; Martelloni et al. 2012).

There are few typical problems in the definition of the rainfall-threshold curves. First, the rainfall actually responsible for the triggering of landslides is frequently unknown. In fact, the exact timing of occurrence is usually undefined (Crosta and Frattini 2001; Aleotti 2004; Staley et al. 2013) and the rainfall is usually measured at rain gauges that may be kilometres apart from the landslide location. For the latter, a major improvement can be made by using meteorological radar and satellite data (David-Novak et al. 2004; Chiang and Chang 2009; Marra et al. 2014; Nikolopoulos et al. 2015; Iadanza et al. 2016; Postance et al. 2018; Mathew et al. 2014; Brunetti et al. 2018). In particular, rainfall radar data provide a good picture of the rainfall pattern in space, but the estimates of rainfall are burdened with errors that are very often quite significant (Krajewski 1987; Schleiss et al. 2020). The sources of errors in radar rainfall measurement can be categorised as: (i) errors in estimating radar reflectivity factor, (ii) variations in the Z-R relationship, and (iii) gauge and radar sampling differences (Hitschfeld and Bordan 1954; Wilson and Brandes 1979).

A second problem with rainfall threshold is that landslide triggering is controlled by site-specific variables, such as local topographical and soil conditions. Although local rainfall thresholds may better characterise the actual site-specific triggering condition (Crosta 1998), most of the rainfall thresholds are defined at regional scale, for areas of similar meteorological, climatic, and physiographic characteristics (Guzzetti et al. 2007) for two reasons. First, because the need for a landslide database large enough to be statistically significant usually requires the information to be generalised over larger areas. Second, because regional-scale thresholds are more suited for operational landslide warning systems (Guzzetti et al. 2007). At the same time, the same amount of rainfall in the same location may trigger a landslide or not, according to other factors that may change in time such as initial soil wetness (Godt et al. 2006; Mirus et al. 2018; Marino et al. 2020), root cohesion (Wu et al. 1979; Schmidt et al. 2001), frost conditions (McRoberts and Morgenstern 1974). Therefore rainfall thresholds should be probabilistic in nature (Frattini et al. 2009; Berti et al. 2012), with different level of rainfall and triggering probability

potentially associated to different warning actions (Mirus et al. 2018; Piciullo et al. 2017). The modelling of probabilistic thresholds requires the identification of non-triggering events (Berti et al. 2012), i.e. strong rainfall events that do not trigger landslides. This identification is very difficult and somehow arbitrary because it requires a minimum level of rainfall to be used for the selection of rainfall events that do not cause failure. On the other side, it may be extremely important because it strongly affects the threshold levels.

A third problem with rainfall thresholds is the evaluation of their quality. Several performance techniques have been used, such as threat score (TS) and True Skill Statistic (TSS), but the most common approach is to use receiver operating characteristic (ROC) metrics. The ROC analysis has been frequently adopted to assess the success of binary prediction in different fields, such as medical testing (Goodenough 1974, Hanley and McNeil 1982), machine learning (Egan 1975, Adams and Hand 1999), and also landslide studies (Beguieria 2006; Frattini et al. 2010). In the definition of rainfall triggering thresholds, ROC curves have been recently adopted for evaluating the performance of the thresholds (Staley et al. 2013; Mathew et al. 2014; Gariano et al. 2015; Piciullo et al. 2017; Leonarduzzi et al. 2017; Hong et al. 2018; Vaz et al. 2018), but also as a technique to identify the optimal threshold that minimise the false alarms while maximising the true alarms (e.g. correct alarms) (Postance et al. 2018).

A problem that is widely neglected in the definition of rainfall thresholds to be used as a practical tool for early warning is that issuing an alert always produces economic consequences. A correct classification of a harmful rainfall event (true alarm) with an alert issued when the threshold is exceeded may result in a timely evacuation of people from endangered locations, saving lives and goods, but also causes the interruption of roads or economic activities, with an indirect cost that may be relevant. The misclassification of rainfall events also produces economic costs that may be significantly different in case of false alarm or missed alarm. Hence, the performance of the thresholds could be evaluated by assessing these costs, in order to select the best thresholds, i.e. the one that minimises costs to society. This has been typically done in disciplines such as machine learning (Drummond and Holte 2000; Provost and Fawcett 2001) and biometrics (Briggs and Ruppert 2005), but also for landslide susceptibility (Frattini et al. 2010).

As far as we know, the techniques used in the literature to define rainfall thresholds do not account for misclassification costs. This limitation is significant as the costs of misclassifications are very different depending on the error type. Error type II (missed alarm) means that a harmful rainfall event is classified as non-harmful, and the alert is not issued. This exposes people to landslide hazard, potentially causing loss of lives and damages to mobile goods, such as cars. The missed alarm misclassification cost, $C(-|+)$, is equal to the loss of elements at risk that can be impacted by landslides during the event. This cost is aleatory, since the landslides may occur in non-inhabited slopes, and depends on the economic value and the vulnerability of elements at risk (e.g. lives, buildings and lifelines), the intensity of landslides, and the spatial distribution of landslides with respect to the elements at risk. In general, we can say that these costs are extremely difficult to quantify precisely, due to the uncertainty about the variables that control such costs.

Error Type I (false alarm) means that a non-harmful rainfall event is classified as harmful, thus causing the issue of an alert. Hence, the false alarm misclassification cost, $C(+|-)$, amounts to the indirect costs related to evacuation, interruption of infrastructures, and suspension of economic activities. This cost is certain and may be potentially calculated as a function of the socio-economic condition of the area. Based on these costs, the optimal threshold may change significantly, as illustrated in this paper.

The research hypothesis of this paper is that classification costs accounted for the selection of the optimal thresholds to be used for landslide risk management. We tried to demonstrate this hypothesis by developing new ROC-based thresholds in Lombardy Region (Northern Italy). For this, we needed to face several problems, such as the definition of a representative rainfall value, the selection of non-triggering events and the quantification of reliable misclassification costs.

Methods

In general, the definition of rainfall thresholds requires: (i) a statistical or empirical approach to define the triggering thresholds and their probability; (ii) an inventory of landslides that were triggered during a storm, with a reliable idea of the triggering time; and (iii) rainfall data of enough resolution that can be correlated to landslide triggering or non-triggering. Further, the selection of an optimal minimum-cost threshold requires the estimation of the probability of landslide events, and the quantification of misclassification costs.

The cost-sensitive rainfall threshold

To explicitly represent costs in the definition of the triggering rainfall thresholds, we adopted the cost curve approach (Drummond and Holte 2006; Frattini et al. 2010) (Fig. 1b). The cost curve represents the normalized expected cost (NEC) as a function of a probability-cost (PC(+)) function.

The normalized expected cost, NEC is calculated as:

$$NEC = \frac{(1-TAr) \cdot p(+)\cdot C(-|+) + FAR \cdot p(-)\cdot C(+|-)}{p(+)\cdot C(-|+) + p(-)\cdot C(+|-)} \quad (1)$$

where TAr is the true alarm rate, FAR is the false alarm rate, $p(+)$ is the a priori probability of having a landslide, $p(-)$ of not having, $C(-|+)$ are the misclassification costs associated to missed alarms, and $C(+|-)$ to false alarm. The expected cost is normalized by the maximum expected cost that occurs when all cases are incorrectly classified, i.e. when FAR (false alarm rate) and MAr (missed alarm rate) are both one. The maximum normalized cost is 1 and the minimum is 0.

The probability-cost function, $PC(+)$, is:

$$PC(+) = \frac{p(+)\cdot C(-|+)}{p(+)\cdot C(-|+) + p(-)\cdot C(+|-)} \quad (2)$$

which represents the normalized version of $p(+)\cdot C(-|+)$, so that $PC(+)$ ranges from 0 to 1. When misclassification costs are equal, $PC(+)=p(+)$. In general, $PC(+)=0$ occurs when cost is only due to negative cases, i.e. positive cases never occur ($p(+)=0$) or their

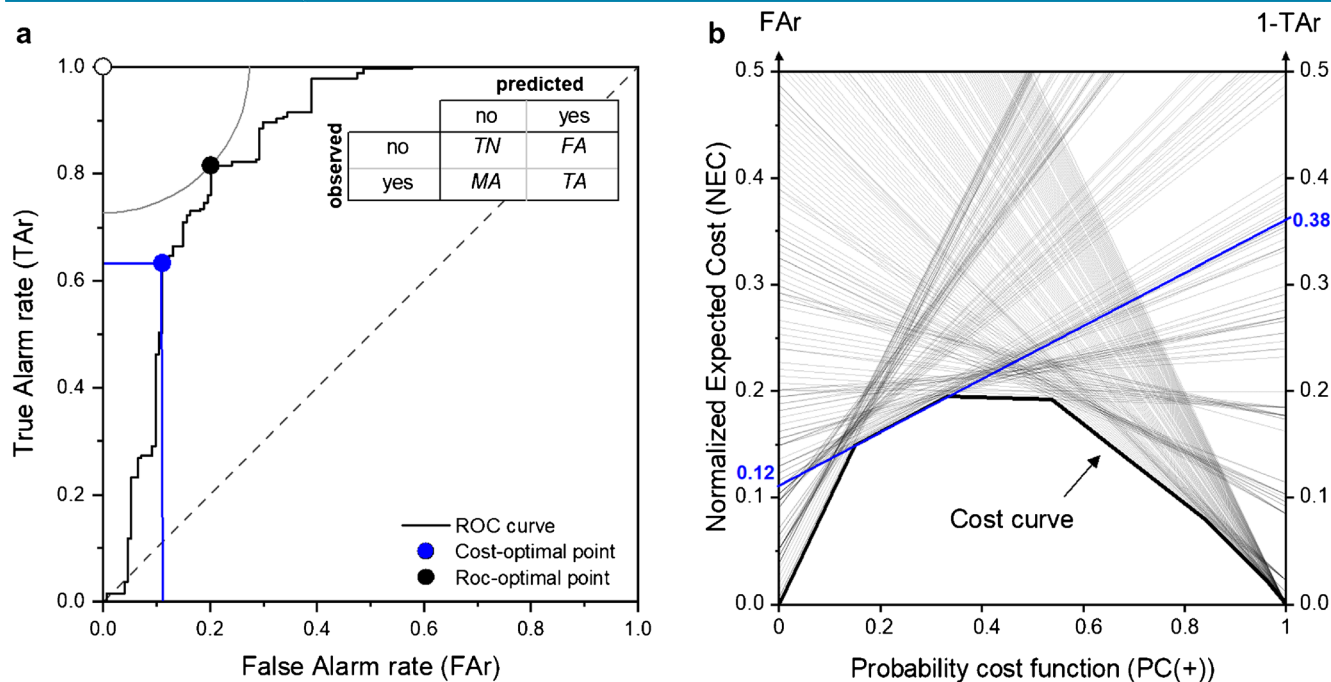


Fig. 1 ROC (a) and cost-curve (b) approaches. In the ROC space (a), each point of the curve represents a rainfall cutoff with a certain value of True Alarm rate ($TAr = TA / (TA + MA)$) and False Alarm rate ($FAR = FA / (FA + TN)$): the best cutoff is the one closest to the upper left corner of the graph. In the cost-curve space (b), a single rainfall cut-off value, which would be a single point in ROC space, is a straight line representing the normalized expected cost (NEC) as a function of the Probability cost function ($PC(+)$): for a certain value of $PC(+)$, the best rainfall cutoff corresponds to the lowest line. The cost curve is the envelope of the best cut-off lines

misclassification cost, $C(-|+)$, is null. $PC(+)$ = 1 corresponds to the other extreme, i.e. $p(-) = 0$ or $C(+|-) = 0$.

The optimal cost-sensitive cutoff is therefore the one that minimises the normalized expected cost, given a certain probability-cost function.

In order to calculate these cut-off values, it is therefore necessary to define:

- i) The a priori probability of having a landslide, $p(+)$ or no landslide, $p(-) = 1 - p(+)$ for the rainfall events considered in the analysis (i.e. events that overcome the threshold used in Lombardy Region for issuing the second-level meteorological alert);
- ii) The costs of misclassification of the different error types, $C(+|-)$ and $C(-|+)$.

The misclassification costs depend on landslide magnitude, extension, and organisational model needed to deal with. The main costs are the direct costs for buildings and infrastructures damages, for evacuation, for civil protection engagement, and for human costs (both killed and injured people), and the indirect costs for interrupted traffic and for unrealized gain of evacuated people. Among these costs, we considered only those that occur due to misclassification (missed alarms, and false alarms) and may be avoided in case of opposite decision. The direct costs of damage to buildings and infrastructures were not considered because they would occur, in case of landslide, whatever the fact that we alert or not.

When the costs of misclassification and the a priori probability of having a landslide, $p(+)$ or no landslide, $p(-)$ are equal, the cost-curve approach reduces to Receiver Operating Characteristic (ROC) analysis, which only maximises the probability of detection and minimises false alarm rate (Fig. 1a).

For the construction of the triggering rainfall threshold curves, the cost-curve approach or the ROC analysis are applied to different rainfall durations (D) to identify the rainfall-intensity (I) cut-off points in the I - D plot. Therefore, the cut-off points are interpolated by the least-square regression.

Demonstration site

The cost-based approach is demonstrated in the central sector of Alpine area of Lombardy Region (Italy), which extends about 350,000 km², ranging in elevation from about 200 m to 3050 m a.s.l. (*Pizzo di Coca*), where hundreds of landslides were triggered during four extreme rainfall events in June 1997, November 2000, November 2002, and July 2008 (Fig. 2).

This area is located in the Central European Alps and is characterised by three main structural units: Southern Alps, Penninic unit, and Austroalpine domain. The Penninic and the Austroalpine domains are located to the north, respectively in the eastern and western parts. The first represents the deepest part of the Alpine belt and is composed of metamorphic rocks from oceanic lithosphere and European margin basement. The second one includes metamorphic and sedimentary formations detached from the lithosphere during the orogenesis. The Southern Alps domain is separated from the other domains by the Insubric fault zone (Fig. 2), a regional trending east-west fault. This domain

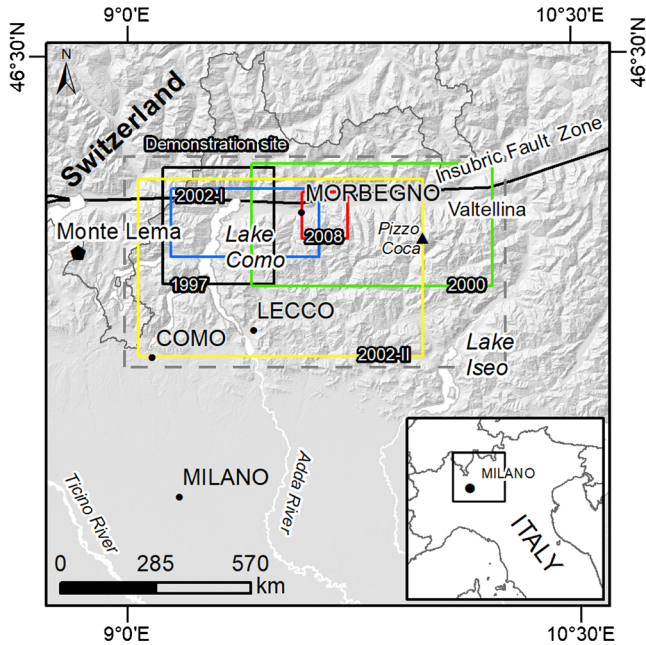


Fig. 2 Demonstration area. The coloured boxes show the extent of the areas affected by the landslide events considered in the paper. Black: June 1997 event; green: November 2000 event; blue: November 13–16 2002; yellow: November 23–26 2002; red: July 2008

includes a fold-and-thrust system characterised by basement and sedimentary cover rocks, and two younger intrusive bodies, the Adamello and the Val Masino-Bregaglia. The geological structure, the lithologies of different strength and the climate result in a high relief, deep valleys and high mountains. Most of the demonstration site has a continental climate with rainfall concentrated in spring and autumn, ranging between 950 and 2000 mm/a.

From the landslide database of the Lombardy Region (GeoIFFI, Frattini et al. 2003; Trigila et al. 2010), which includes 144,000 landslides classified as rockfalls, debris flows, shallow landslides, and deep-seated landslides, we extracted shallow landslides that occurred during the 1997, 2000, 2002 and 2008 rainfall events. Then, we improved the inventory by mapping additional landslides from digital orthophotos, obtaining 607 landslides (271 in 1997, 194 in 2000, 86 in 2002, and 56 in 2008). A careful historical research, based on bibliographic and archival data source, allowed the definition of the day and, in some cases, the hour of the occurrence of most of the landslides.

Both rain-gauge and radar data were collected to extract triggering and non-triggering rainfall events. Rain gauge data were collected from ARPA Lombardia (<https://www.arpalombardia.it>). According to the data availability at the time of the event, the analysis was performed on 22 rain gauges for the June 1997 event, 33 for the November 2000 event, 34 for the November 2002 event, and 99 for the July 2008 event. Radar data were collected from Monte Lema radar images, provided by Meteo Swiss. The Monte Lema radar station, installed in 1993, is located at an altitude of 1626 m a.s.l. and it operates with a sampling interval of 5 minutes. It is a C-band Doppler radar with an antenna diameter of 4.2 m and it covers a cylindrical volume with a radius of 200 km and a height of 12 km; the full volume is scanned with a 1° beam at 20

elevations. The spatially distributed radar data consists of GIF images representing rainfall intensities in mm/h for 5-minutes time intervals. Each image consists of 1 km × 1 km resolution pixels.

Early warning management

The alert system adopted by Lombardy Region is well-structured in two phases: forecasting, based on weather models, and monitoring, to integrate direct and instrumental observations. The system is based on three alert levels: ordinary (1st level), moderate (2nd level), and high (3rd level), associated with rainfall thresholds. The ordinary alert does not activate specific actions but the notice of attention to regional and municipal authorities. The moderate alert activates the civil protection operational centres and the surveillance points, while the high alert activates the alarm operational phase, including ban on risk areas, road closure, evacuation of the population, and rescue of people in danger. The adoptable actions activated at different alert levels are not mandatory. The responsibility of the civil protection managers encloses a case-specific discretionary assessment of the actions to be taken, even to limit crying-wolf effect (Brenitz, 1984) in terms of loss of confidence in the warning system.

Analysis and results

Radar data correction

In this research, the systematic comparison of rain-gauge and radar data showed a substantial difference between the two sources. To combine the advantages of accurate point rainfall estimates of rain gauges with the large areas survey ability of radar and to improve the accuracy of radar while preserving its spatial description of rainfall fields, we applied a radar bias adjustment method. First of all, we calculated for each rain gauge station and for each rainfall event a bias adjustment factor, β , defined as:

$$\beta = \frac{1}{N} \sum_{i=1}^N \frac{Rg_i}{Rr_i} \quad (3)$$

where Rg is the rain-gauge data, Rr is the radar data at the location of the rain gauge, and N is the number of hourly rainfall data. Values of β greater than 1 imply a radar underestimation compared to the rain gauges and values between 0 and 1 imply a radar overestimation.

In order to attain the best radar correction, three different approaches were performed: i) a mean-field bias adjustment (Smith and Krajewski 1991; Anagnostou et al. 1998; Seo et al. 1999; Ochoa-Rodriguez et al. 2019), consisting in the calculation of an average regional value of bias adjustment factor, ii) a correlation function between bias adjustments and geographic parameters, such as the distance from the radar, the elevation, and the visibility in order to account and correct limitations of the instrument, and the potential horizontal and vertical errors (Gabella and Notarpietro 2004), iii) an interpolation of the local bias adjustments by using the inverse distance weighting (IDW) interpolator, in order to adapt the radar data to the rain gauge data, still maintaining the fine granularity of radar information

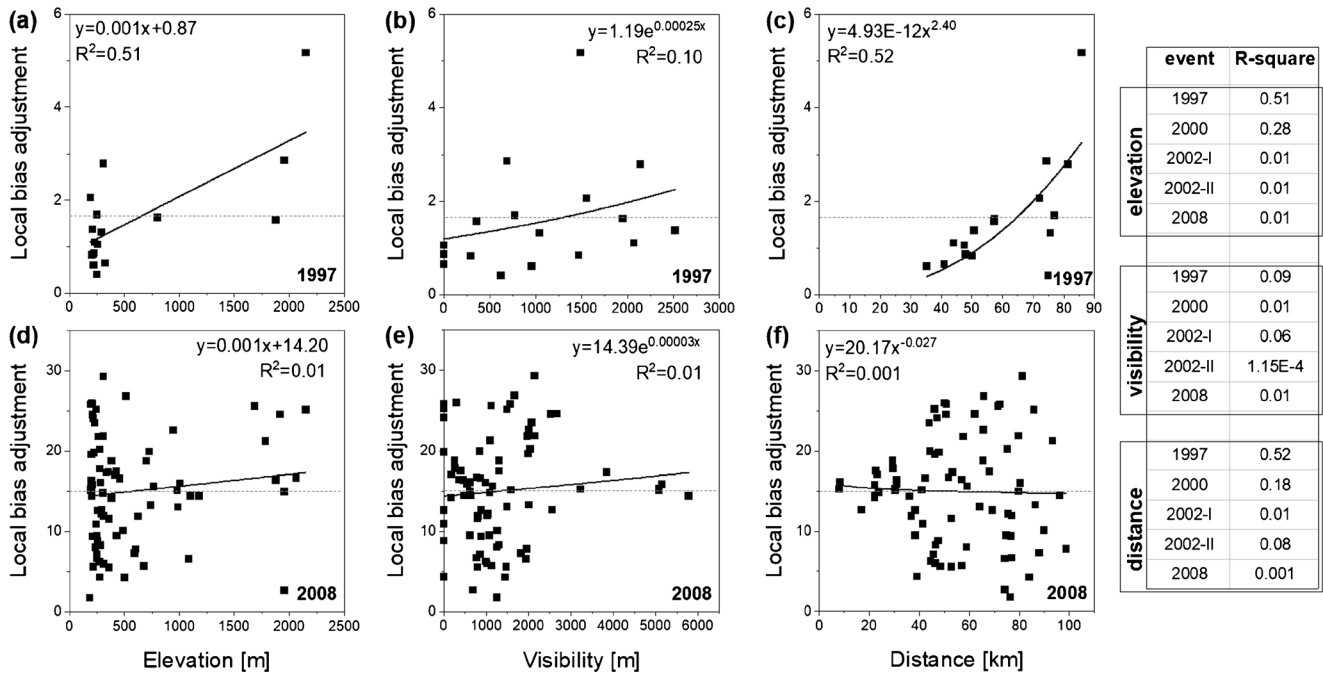


Fig. 3 Relationships between the bias adjustment factor and the elevation of the rain-gauge stations (a and d), the visibility (b and e), and the distance from the radar (c and f) for two rainfall events: a–c) June 1997; d–f) July 2008. The black solid line is the regression line, whose function and R2 is reported in each figure. The dashed line is the average adjustment factor of each event

The mean-field bias adjustment (β , Eq. 3) was applied considering rain gauges within 100 km of distance from Monte Lema radar station. The resulting precipitation mean-field bias adjustment values in many cases overrate local bias adjustment values, leading to great errors in the rainfall radar estimate. For this reason, this strategy was abandoned.

To compensate the radar underestimation bias in the overall area, we tried to correlate the bias adjustment factor with: (i) distance between the radar and the gauges, (ii) the elevation of the gauges, and (iii) their visibility (the minimum height that needs to be added to each rain gauge to make it visible by the radar) (Fig. 3). While in some cases we could identify significant trends (e.g. with distance in 1997 event, Fig. 3c), these trends were not consistent for the different events, giving low R-square values in the least-square regression. Given the impossibility to identify consistent relationships, we abandoned also this strategy.

The interpolation of the bias adjustment factors has been developed by applying the inverse distance weighting (IDW) exact interpolator. Figure 4 reports the interpolated maps of the local bias adjustments. In 1997, 2000 and 2002 event I and event II (A, B, C and D respectively), the IDW interpolation displays an increase in the local adjustments in the north-east direction, more evident in the 2000 event (B), whereas in the 2008 event a minor range of bias interpolation is reported. The values of bias adjustment are almost always larger than 1, suggesting that the rain-gauge values are always greater than the radar measurements (Fig. 5), probably due to a decreasing vertical profile of reflectivity with height, combined with beam shielding and/or occultation by orography (Gabella and Notarpietro 2004).

The values of interpolated adjustment factors were used to correct rainfall radar estimations for each rainfall event (Fig. 6),

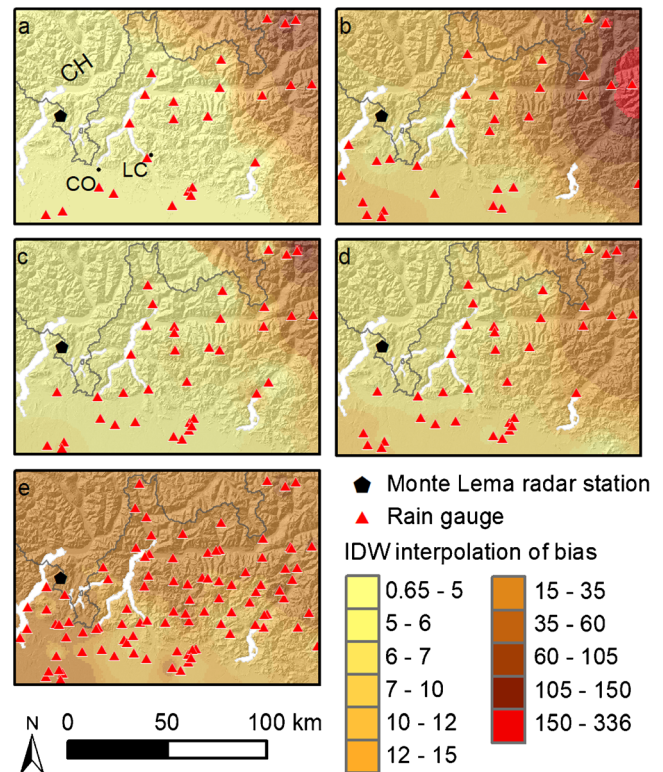


Fig. 4 Interpolation of bias adjustment factor by means of inverse distance weighting: a) June 1997; b) November 2000; c) November 13–16, 2002; d) November 23–26, 2002; e) July 2008. In (a), CO = Como; LC = Lecco; CH = Switzerland

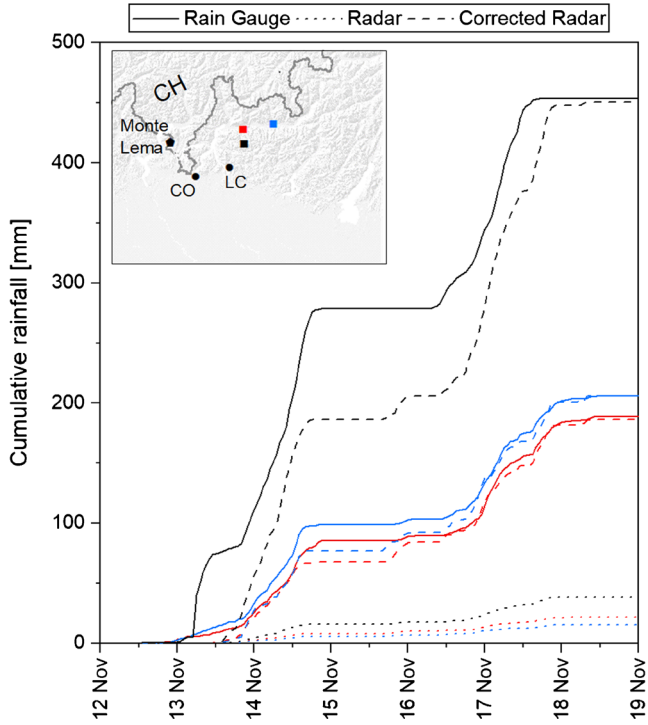


Fig. 5 Example of radar correction for three rain gauges for the 2002 rainfall event. Correction is made by using the Interpolated bias adjustment factors (see Fig. 4)

multiplying the interpolated map of the local bias adjustments by the radar map.

Triggering and non-triggering rainfall data

The dataset of non-triggering rainfall (NLR) values was obtained by analysing hourly rainfall data of 14 rain gauges evenly distributed over Lombardy Region from 2008 and 2017 (years). For each station, we extracted the maximum monthly I-D values for the same durations of the triggering events; we removed solid precipitations occurred in winter and precipitations related to all landslide events reported in archival data source. Antecedent soil moisture conditions were not considered. In order to filter out very low rainfall values that may be affected by measurement uncertainty, and that will never trigger landslides due to the negligible infiltration into the soil, we selected as non-triggering rainfall (NLR) values for the statistical analysis only those that exceed a threshold used in the Lombardy Region for issuing the second-level meteorological alert, i.e. rainfall I-D values with a return time of 2 years. The second-level alert was chosen for the analysis because it corresponds to a level of pre-alarm in view of the third-level alert, for which an evacuation could be required. By using this filtering strategy, a total of 332 NLR values were collected.

The landslide-triggering rainfall (LR) values were extracted for each landslide occurred during the five events (1997, 2000, 2002-I, 2002-II, 2008) by using the local bias-corrected radar data. In lack of the exact information of the triggering time for each single landslide occurred during an event, the rainfall duration was assumed to be equal for each landslide and defined from the

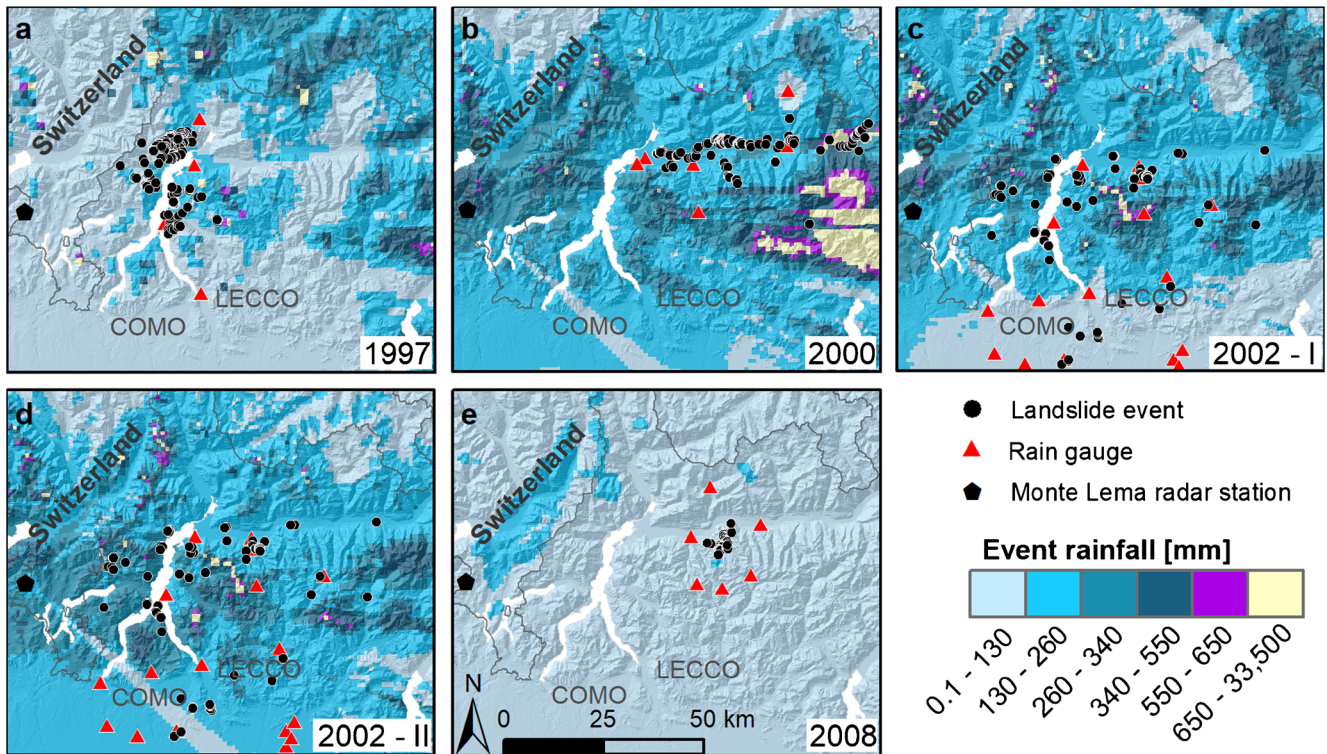


Fig. 6 Study area and rainfall events measured by Mt Lema weather radar, by using corrected radar data: a) June 27 and 28 (event duration: 27 h), 1997; b) November 12–18, 2000 (124 h); c) November 13–16, 2002 (79 h); d) November 23–26, 2002 (72 h); e) July 12–13, 2008 (35 h)

beginning of the rainfall to the known time of occurrence of most of the landslides, as obtained from archival sources (Fig. 7). The 2000 and 2008 rainfall were considered single events because landslides occurred only at the end of the last rainfall peak. A total of 607 landslide-triggering rainfall (LR) values were detected.

Cost-sensitive triggering rainfall thresholds

Cost-sensitive thresholds require the definition of an a priori probability of having a landslide and the costs of different error-type misclassifications.

The a priori probability of landslide occurrence was calculated by dividing the annual frequency of rainfall events causing at least one documented landslide by the annual frequency of rainfall events with a return period exceeding two years (i.e. the second-level meteorological alert threshold in Lombardy Region), which is the same filter value used to select the non-triggering rainfalls. Rainfall events causing landslides were detected from a detailed analysis of historical documents, landslide inventories and specific

Google Search for the period 2010–2019. Landslide events occurred before 2009 were not considered because the catalogue was not complete, apart for large events such as the ones that we used for building the rainfall threshold curves. A total of 11 events were collected. On the other hand, 42 rainfall events with a return period larger than two years were recorded by the 14 rain gauges located in the demonstration area for the period 2008–2017. The resulting $p(+)$ and $p(-)$ amount to 0.26 and 0.74, respectively.

To assess the misclassification costs, we made reference to a representative emergency scenario in the Sondrio province that involves a mountain village located in a landslide-prone area (Civo village, close to Morbegno town, Fig. 2). For this scenario we assumed consequences both in case of alarm and in case of missed alarm, and we quantified these consequences based on past events occurred in Lombardy Region and the literature (e.g. Guzzetti et al. 2005, for human costs in terms of casualties and injuries).

In case of alarm, the scenario is assumed to cause a 1-day evacuation of 200 people, the interruption of a local road

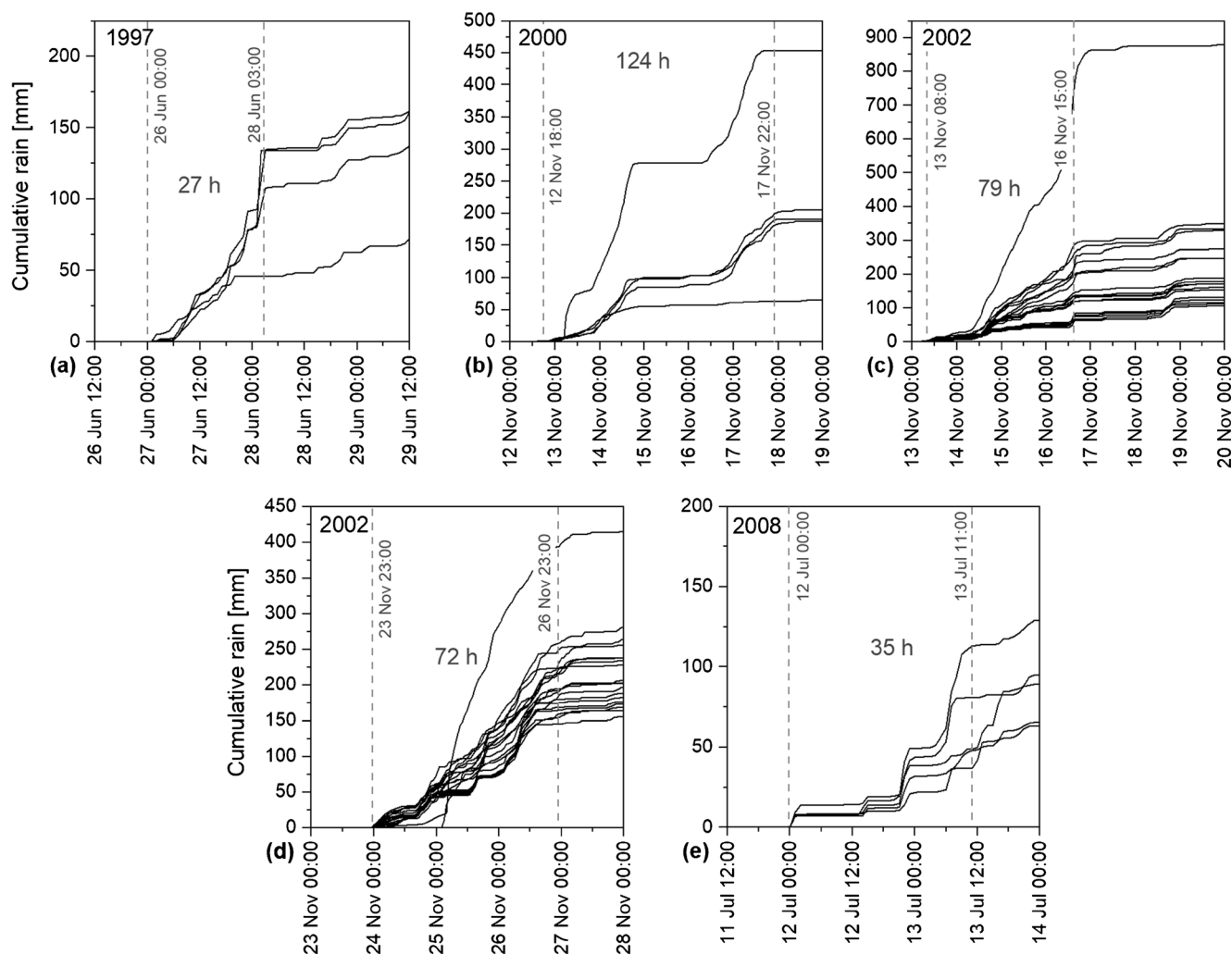


Fig. 7 Cumulative rainfall recorded at rain gauges during the five rainfall events. Although the figures show that the time of rainfall peaks may be shifted due to the storm motion across the mountains terrain, a constant time-window was selected for each event due to the lack of information about the exact timing of landslides occurrence

Table 1 Misclassification costs associated to false alarm (FA) and missed alarm (MA). The unit cost for casualties includes medical costs (1965€) and an average unrealized gain of 940,291€ (Italian Ministry of Infrastructure and Transport 2012). The unit cost for injured people is calculated as a percentage of casualty cost, according to European Conference of Ministers of Transport (2002), corresponding to 13% and 1% of total casualty cost (also including moral cost of 561,734€). Heavily injured corresponds to cases when the hospital time of attendance is greater than 24 h (European Injury Data Base)

Cost item	Unit cost (€/person day)	Number of involved people	Scenario cost (€)	Source for unit cost
Civil protection (FA)	500	100	50,000	Large civil protection exercise in 1990 (http://archivio.corriere.it/Archivio/pro/view.shtml#/MzovZXmVaXQvcmNZZGF0a59ANTY40T1%3D)
Evacuation (FA)	55	200	11,000	Government contribution during the last emergency (e.g. Deliberazione Giunta Regionale n. 582 del 12 ottobre 2009, Regione Abruzzo)
Loss of profit for evacuees (FA)	130	65*	8450	Per capita nominal GDP for Sondrio Province, calculated based on 200 working day (source: Italian Institute of Statistics, ISTAT)
Loss of profit for commuters by car (FA)	130	135^	17,550	Per capita nominal GDP for Sondrio Province, calculated based on 200 working day (source: Italian Institute of Statistics, ISTAT)
Casualties (MA)	942,256	0.50	471,128	Italian Ministry of Infrastructure and Transport 2012
Heavily injured (MA)	197,228	0.09	17,751	Italian Ministry of Infrastructure and Transport 2012
Slightly injured (MA)	16,985	0.01	170	Italian Ministry of Infrastructure and Transport 2012

*200 evacuees; 65% in working age between 20 and 64 (source: Italian Institute of Statistics, ISTAT); 50% unable to perform smart working (Source: Italian Institute of Statistics, ISTAT): $200 \times 0.65 \times 0.5 = 65$

^270 commuters by car (source: traffic data from Regione Lombardia along local road Civo-Morbegno); 50% unable to perform smart working (source: Italian Institute of Statistics, ISTAT): $270 \times 0.5 = 135$

connecting Civo to Morbegno used an average by 270 commuters by car, the activation of 100 Civil Protection technicians and volunteers, the loss of profit for both evacuees and commuters.

In case of missed alarm, the scenario is assumed to cause 0.6 affected people, calculated from the average number of people affected during the 1997, 2000, 2002, and 2008 events. Of these, 83.2% are expected to die, 15.1 to be heavily injured, and only 1.7% lightly injured (Guzzetti et al. 2005), totaling 0.50, 0.09 and 0.01 people per event, respectively. Costs related to vehicles and other moveable assets were neglected.

Table 1 summarises the expected misclassification costs.

Based on the misclassification costs and the a priori probability calculated for the reference scenario, we applied the cost curve approach to identify the optimal intensity threshold for each duration, and we interpolated the threshold to obtain the cost-sensitive threshold curve ($R^2 = 0.86$):

$$I = 50.24 \times D^{-0.77} \quad (4)$$

For comparison, we also built ROC-based rainfall threshold curves by using either the corrected radar data and nearest rain-gauge rainfall data (Fig. 8).

The normalized expected cost (NEC) for the cost curve threshold results 0.25, while the NEC value of the corresponding ROC threshold with radar data, obtained performing a cost curve analysis taking into account equal cost and equal probability of $p(-)$

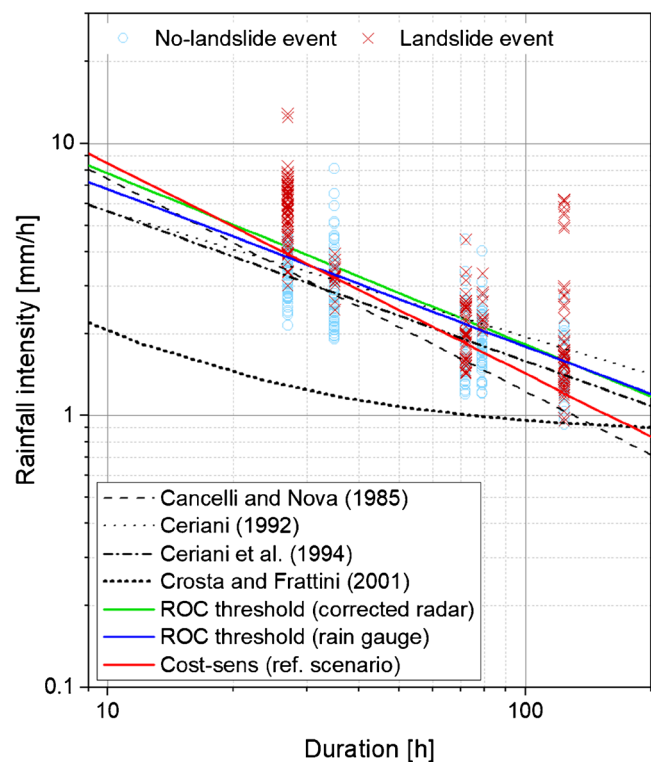


Fig. 8 Landslide-triggering rainfall threshold curves based on ROC approach and cost-curve approach and comparison with literature thresholds proposed for Lombardy Region: Cancelli and Nova 1985 ($I = 44.668 \times D - 0.78$); Ceriani, 1992 ($I = 16.24 \times D - 0.46$); Ceriani et al. 1994 ($I = 20 \times D - 0.55$); Crosta and Frattini 2001 ($I = 12 \times (1/D + 0.07)$)

and $p(+)$, is 0.30. Therefore, in terms of avoidable cost, the new cost-sensitive threshold outperforms the ROC threshold.

Discussion

Cost-sensitive rainfall thresholds

Cost-sensitive rainfall thresholds allow to account for misclassification costs in the definition of the thresholds. Working on a reasonable reference scenario, we were able to produce a curve that minimises the normalized expected cost with respect to the ROC-based approach. The cost-sensitive threshold curve lies below the ROC-based curve, especially for longer durations, because the costs associated to missed alarms are much higher than the costs for false alarm (Fig. 8). Given these misclassification costs, it is reasonable to expect that an optimal cost-sensitive threshold curve should prefer to issue more false alarm in order to minimise missed alarm costs with respect to a threshold curve in which the costs are assumed to be equal.

Due to the uncertainty in the definition of the reference scenario, we performed a sensitivity analysis to assess the weight of the controlling parameters on the threshold. Starting from the reference scenario, we varied one at a time the following parameters, multiplying then by 0.2, 0.3, 0.7, 1.5, 3, and 5: days of evacuation, evacuated people, involved emergency technicians, evacuated workers, involved commuters, number of fatalities, and number of injured people. In total, we calculated 42 scenarios for each rainfall duration. The unit costs were included in this sensitivity analysis because they are homogenous in the entire country and they remain constant for any possible alert scenario.

Figure 9 illustrates the results for the rainfall duration of 27h (see Supplementary F1 for other durations). We can identify two different behaviours for parameters that control the MA misclassification costs (number of deaths and injured people) and parameters that control the FA misclassification costs (days of evacuation, evacuated people, involved emergency technicians, evacuated workers, involved commuters).

An increase in the number of deaths or injured people causes an increase of MA misclassification costs. Therefore, the optimal cutoff needs to reduce the MA rate with respect to the FA rate, leading to an increase of the FAR/MAR ratio (Fig. 9a) and a decrease of the rainfall cutoff (Fig. 9b). On the contrary, an increase in parameters that control the FA misclassification costs leads to a decrease of the FAR/MAR ratio (Fig. 9a), which is obtained with an increase of the rainfall cutoff (Fig. 9b).

In particular, we can observe that for the rainfall duration of 27 h the number of fatalities and the number of evacuation days are the most sensitive parameters. On the other side, the number of injured people, involved commuters, evacuees, emergency technicians and evacuated workers are almost insensitive. The parameters that involve higher costs are most sensitive. The fatalities item has the greater unit cost, while the evacuation includes all the false alarm costs because it involves that all the alarm actions are carried out.

False-alarm cost varies for different rainfall duration (as shown in the Supplementary Material), and this could affect the shape of the threshold that shall be flatter and no more log-linear. So, when the costs are considered, the threshold equation may be different, also considering that costs may vary with the duration. We

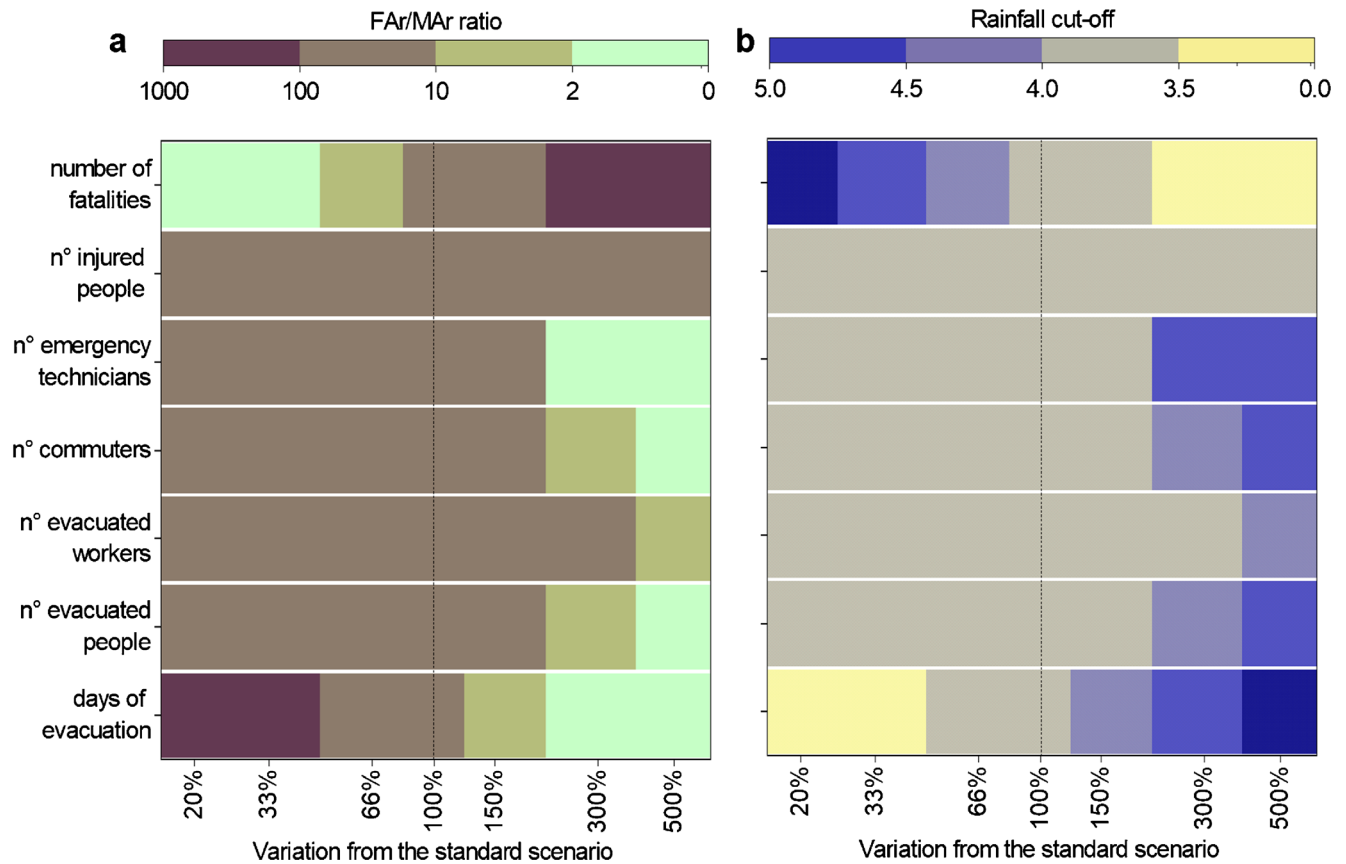


Fig. 9 Sensitivity of (a) the FAR/MAR ratio and (b) the rainfall cutoff (in mm) to changes of parameters that control the misclassification costs for a rainfall duration of 27 h. FAR/MAR ratio represents how much is advantageous to produce false alarms with respect to missed alarms

preferred to adopt the standard log-linear ID curve to keep the focus on how each cost item affects the cut-off definition regardless the rainfall duration.

Based on the range of variation of costs used for the sensitivity analysis, we defined three different threshold-scenarios, in addition to the reference scenario, in order to evaluate the importance of the misclassification costs in the rainfall threshold curves:

- A “maximum missed alarm costs scenario” with minimum false alarm cost: $c(-|+)_{MAX} c(+|-)_{MIN}$;
- A “maximum false alarm cost scenario” characterised by minimum missed alarm cost: $c(-|+)_{MIN} c(+|-)_{MAX}$;
- An *equal costs* scenario characterised by same misclassification costs both for missed-alarm and false-alarm costs. This scenario differs from the ROC-based threshold because it accounts for the probability of landslide events. If the probability of rainfall events were 0.5 for all the durations, this scenario would be identical to the ROC-threshold.

Figure 10 shows the effect of different cost scenarios on the rainfall threshold curve. The maximum missed alarm costs scenario threshold curve is the lowest with respect to all the other scenarios. This is because an increase of missed alarm costs would favour a low value of the rainfall threshold curve, which therefore

would issue continuous alerts (with numerous false alarms) in order to avoid missed alarms. In such scenario, in fact, the human costs of missed alarms are more than 6 times higher than Civil Protection, evacuation and traffic interruption costs.

On the other side, the maximum false alarm cost scenario threshold curve is higher than the reference scenario cost-sensitive threshold curve and much higher than the equal-cost scenario and the ROC-based threshold curves. Taking into account the greater probability of $p(-)$ with respect to $p(+)$ in the investigated area, and since the avoidable costs of missed alarm scenario are part of an expensive bet (amounting to potential fatalities and/or injuries) whereas in false alarm they are certain (amounting to the indirect costs related to evacuation, interruption of infrastructures, and suspension of economic activities), preferring a higher threshold that increases the risk of fatalities and injured people is suitable from a cynical pure-cost perspective, even if potentially questionable from an ethical point of view. The major limitation of this decision-making model is that it is reasonable only under rare-events conditions, because in case of frequent events, the choice of adopting a lower threshold is always desirable from a precautionary perspective. In other words, this model can be applied only with a high degree of belief that the probability of events is very low, which is not always the case.

An aspect that was not considered for the thresholds is the crying-wolf effect. Both intuition and theory suggest that false

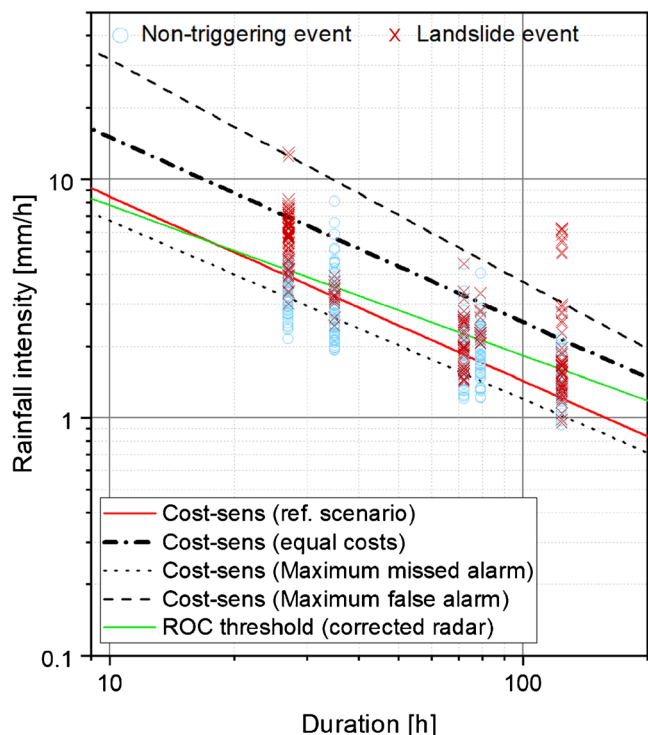


Fig. 10 Landslide-triggering rainfall threshold curves obtained with different cost scenarios. ROC-based curve is also reported for comparison

alarms should reduce warning response (Breznitz 1984; Simmons and Sutter 2009). This may lead to additional costs occurring even in case of true alarm that could potentially affect the optimal cutoff. However, the empirical evidences for the crying-wolf effect are elusive (Dow and Cutter 1998; Sorensen 2000). Moreover, according to the national regulation, the evacuation is managed at the municipal level and carried out by the executive authority, thus limiting the discretionary component linked to the citizen behaviour.

Issues for rainfall threshold construction

The lack of accuracy of both the *rainfall data* and the *inventory of landslides* complicates the definition of an optimal rainfall threshold curve.

While analysing rainfall data obtained by rain gauges and radar, inconsistencies between the two sources were detected. The classical strategy of using the rainfall of the nearest rain gauge provides a rainfall threshold that performs worse than the ones existing in the literature for the same area (Fig. 11). The need for a regionalisation of rainfall data was accomplished by combining the point rainfall estimates of rain gauges with the large areas survey ability of radar, through the interpolation of local bias adjustment values. Unfortunately, a more physically meaning correlation of local bias adjustment with other spatial variables (i.e. elevation, distance from the radar) was not possible because the accuracy and the level of detail of the archive data were not enough to identify significant correlations.

In Fig. 11, we can observe that ROC-based threshold based on corrected radar data outperforms the literature thresholds, with the exception of Ceriani et al. (1992). The new cost-sensitive curve

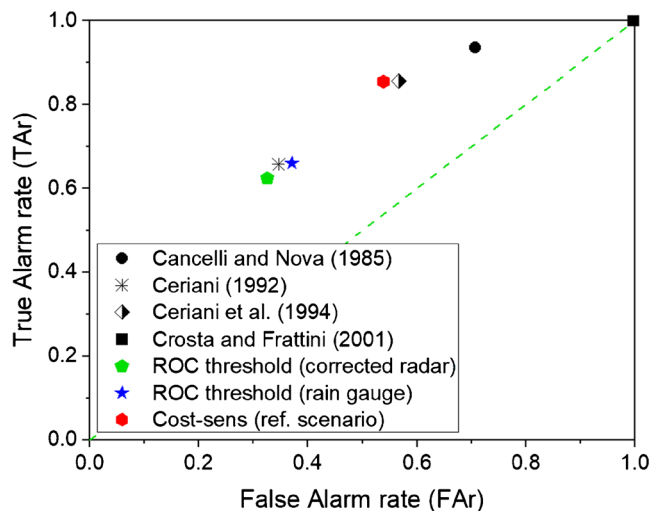


Fig. 11 Performance of ROC-based landslide-triggering rainfall threshold, cost-sensitive threshold and literature thresholds

plots more to the right, because it tends to increase the false alarm rate (Fig. 11). However, these thresholds still provide a performance that is not optimal, as demonstrated by the high value of the false alarm rate, especially for the cost-sensitive curve (Fig. 11). This is due to several factors. First of all, the threshold is defined based on a limited number of landslide-triggering events for which a reliable inventory was available. In addition, even if the location of each landslide is reliable due to an accurate mapping of the events, the exact triggering time is unknown, leading to large uncertainties in triggering rainfall at the landslide site. More in detail, the lack of time information leads to a simplification: the specific critical rainfall of each event is generalised, assuming that all landslides were occurring simultaneously. Moreover, the construction of rainfall thresholds ignores the role of antecedent rainfall and soil moisture, which can both deeply affect the triggering condition at site. In particular, one of the five events (23–26 November 2002) occurred soon after a previous rainfall event, making the initial conditions significantly different from the others. The nature of the rainfall events also affects the rainfall thresholds. The 2008 event was a localised summer cloudburst. For such phenomena, the values of rainfall collected at rain gauges are strongly underestimated (Fig. 12), but also the bias-corrected radar may be slightly underestimated (Marra et al. 2014). This can be observed looking at the landslide-triggering rainfall intensity values (crosses in Fig. 8) for the duration of 35 h (i.e. the 2008 event), which appear significantly lower than the triggering threshold obtained by interpolating all the different durations, giving a high number of missed alarms.

Finally, another key-parameter in the construction of rainfall thresholds is the amount of minimum rainfall that is used to select the non-triggering rainfall (NLR) values, filtering out trivial very-low rainfall, which may bias the analysis by reducing the higher false-alarm rate, leading to lower landslide-triggering rainfall thresholds. To evaluate the effect of this minimum filtering rainfall, we built the threshold curves with four different populations of NLR values: (i) one that consists in the whole dataset of non-

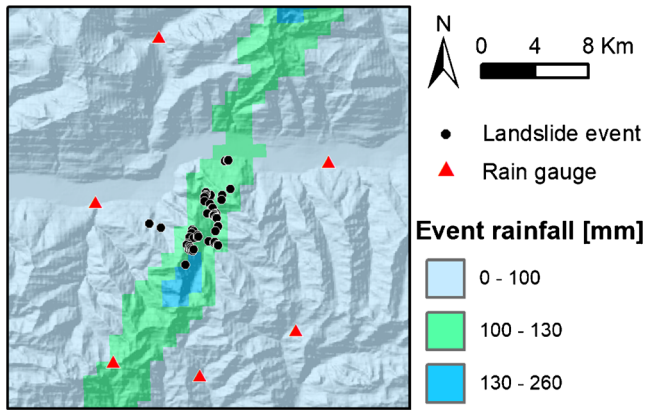


Fig. 12 Detail of the 2008 rainfall event. The cloudburst was extremely localised along a NS strip. Rain gauges mostly did not recognise the rainfall event

triggering rainfall events, (ii) and three derived from filtering the dataset with the Lombardy thresholds adopted by the region civil protection, corresponding to the activation of the 1st (5-year return time), 2nd (2-year return time), and 3rd (1.5-year return time) alert levels. Figure 13 shows that the final threshold curves remain quite stable, thus demonstrating that the cost-curve approach is not much sensitive to the choice of the minimum rainfall. In general, we observe that the lower the minimum filtering rainfall, the lower the threshold curve. Using the whole dataset of non-triggering rainfalls, without any filter, is not convenient because it introduces

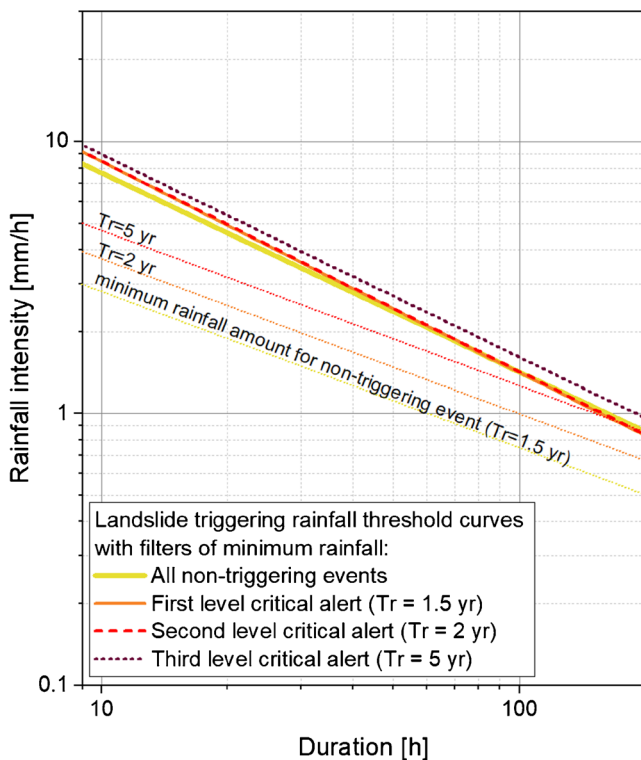


Fig. 13 Landslide-triggering rainfall threshold curves developed by using different amounts of minimum rainfall in the definition of non-triggering events. The minimum amount of rainfall is also reported in the figure

further uncertainty in the cut-off definition by adding trivial very-low rainfall values that may be affected by measurement errors and that give a negligible contribution to the soil strength reduction and to landslide triggering.

Operational implementation of cost-based thresholds

The Lombardy Region alert system works on homogeneous areas, but it is the Civil Protection together with the municipal authority to choose whether and where to implement the possible actions. Experiences from the past showed that evacuation is rare, and limited to single slopes or watershed already known to be prone to landslides. This makes difficult to identify a representative evacuation scenario, which could be a single building, a village, a municipality or a larger area. It is evident that there is no point in formalising each possible evacuation scenario, because both the costs of false and missed alarms will vary locally.

However, although the quantified consequences and costs are site-specific, we believe that the proposed method is applicable at different scales because both false and missed alarm costs mainly depend on the population density. Assuming that both costs varies linearly with the density, we can expect that, within the same socio-economic context and alert system (e.g. northern Lombardy Region), the ratio between the two costs would remain approximately constant by increasing or decreasing the population density. Remaining constant the ratio, the proposed threshold would also remain constant for scenarios with different size.

On the other side, the application of the cost-based approach in a different socio-economic context and with a different alert system requires to recalculate the costs.

Conclusions

The results of this research lead to the following conclusions:

- The new cost-sensitive rainfall thresholds allow to explicitly account for misclassification costs in rainfall threshold definition. This is very important because missed-alarm costs and false-alarm costs may be significantly different, making an increase or decrease of the rainfall thresholds, according to these misclassification costs, convenient. As far as we know, this is the first attempt to build a cost-sensitive rainfall threshold for landslides.
- Considering a representative emergency scenario, we found that the cost-sensitive rainfall threshold curve is lower than the ROC curve (Fig. 8). In fact, since the missed-alarm costs (i.e. human costs for fatalities and injured people) are almost seven times greater than false-alarm costs (587,000 euro and 87,000 euro, respectively), the most convenient threshold should be low enough to minimise the missed alarms (i.e. the number of missed alarms).
- For different socio-economical settings (e.g. different population density, transport facilities, civil protection organisation) the misclassification costs may vary significantly and so their cost-sensitive threshold curves, accordingly. In order to quantify the range of variation of these curves, we made a sensitivity analysis of the main parameters that control costs, and we produced thresholds for two opposite cost scenarios. We find that thresholds range were as much as half an order of magnitude.

- The information regarding rainfall intensity at the landslide site is very sensitive. As reported in the literature, the use of weather radar allows to significantly improve the accuracy of rainfall intensity values at the site of landslide. However, radar data needs a correction of the value that may vary in time and space. In this paper, we applied a simple interpolation of the local bias adjustment to the five landslide-triggering rainfall events, and we did not find more significant correlation with distance from radar or with local morphometric conditions.
- The exact time of landslide occurrence is unknown for most events, especially for large-scale landslide events. Crowdsourced landslide reporting may partially overcome this limitation, but it would be largely ineffective for landslides occurring in remote areas where witnesses are missing.
- The minimum rainfall amount that is used for the definition of a non-triggering rainfall event is commonly an arbitrary choice. In this work we used as minimum rainfall the alert threshold values of Lombardy Region, which are defined in terms of probability of occurrence. However, we showed that the use of different thresholds does not affect significantly the rainfall threshold curves.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10346-021-01707-4>.

Acknowledgements

The authors wish to thank Luca Piciullo for discussions, and Cees Van Westen and Benjamin Mirus for useful and constructive reviews.

Funding

Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement. This work has been supported by Fondazione Cariplo, grant no. 2016-0769 (DARFLIS project). Meteo radar data was provided by Swiss Meteo.

References

Adams NM, Hand DJ (1999) Comparing classifiers when the misallocation costs are uncertain. *Pattern Recogn* 32(7):1139–1147

Aleotti P (2004) A warning system for rainfall-induced shallow failures. *Eng Geol* 73(3–4):247–265

Anagnostou EN, Krajewski WF, Seo DJ, Johnson ER (1998) Mean-field rainfall bias studies for WSR-88D. *J Hydrol Eng* 3(3):149–159

Beguería S (2006) Validation and evaluation of predictive models in hazard assessment and risk management. *Nat Hazards* 37(3):315–329

Berti M, Martina MLV, Franceschini S, Pignone S, Simoni A, Pizziolo M (2012) Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. *J Geophys Res Earth Surf* 117(F4)

Breznitz S, (1984) Cry wolf: The psychology of false alarms. Lawrence Earlbaum Associates, 265 pp

Briggs W, Ruppert D (2005) Assessing the skill of yes/no predictions. *Biometrics* 61(3):799–807

Brunetti MT, Melillo M, Peruccacci S, Ciabatta L, Brocca L (2018) How far are we from the use of satellite rainfall products in landslide forecasting? *Remote Sens Environ* 210:65–75

Caine N (1980) The rainfall intensity duration control of shallow landslides and debris flows. *Geogr Ann* 62:23–27

Campbell RH (1974) Debris flows originating from soil slip during rainstorms in southern California. *Q J Eng Geol London* 7:339–349

Cancelli A, Nova R (1985) Landslides in soil debris cover triggered by rainstorms in Valtellina (central Alps, Italy). *Proc. 4th Int. Conf. on Landslides, Tokyo*

Ceriani M, Lauzi S, Padovan N (1992) Rainfalls and landslides in the alpine area of Lombardia region, central Alps, Italy. In *Internationales symposium: Interpraevent* (pp. 9–20)

Ceriani M, Lauzi S, Padovan N (1994) Rainfall thresholds triggering debris flows in the alpine area of Lombardia Region Central Alps–Italy. I *Convegno internazionale per la protezione e lo sviluppo dell'ambiente montano "Man and Mountain"94*" (pp. 123–139)

Chiang SH, Chang KT (2009) Application of radar data to modeling rainfall-induced landslides. *Geomorphology* 103(3):299–309

Crosta G (1998) Regionalization of rainfall thresholds: an aid to landslide hazard evaluation. *Environ Geol* 35(2–3):131–145

Crosta GB, Frattini P (2001) Rainfall thresholds for triggering soil slips and debris flow. In *Proceeding of 2nd EGS Plinius Conference on Mediterranean Storms, Siena* (Vol. 1, pp. 463–487)

Crozier MJ, Glade T (1999) Frequency and magnitude of landsliding: fundamental research issues. *Z. Geomorph. N.F.* 15:141–155

David-Novak HB, Morin E, Enzel Y (2004) Modern extreme storms and the rainfall thresholds for initiating debris flows on the hyperarid western escarpment of the Dead Sea, Israel. *Geol Soc Am Bull* 116(5–6):718–728

Dow K, Cutter SL (1998) Cry wolf: repeat response to hurricane evacuation orders. *Coast Manag* 26:237–252

Drummond C, Holte RC (2000) Explicitly representing expected cost: an alternative to ROC representation. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 198–207)

Drummond C & Holte RC (2006) Cost curves: An improved method for visualizing classifier performance. *Machine Learning* 65(1):95–130

Egan JP (1975) *Signal detection theory and ROC analysis* "Academic Press. New York

European Conference of Ministers of Transport 2002 Road transport. Road safety. Recent trends in road accidents and policy issues

Frattini P, Crosta GB, Ceriani M, Fossati D (2003) Inventario delle frane e dei dissesti della regione Lombardia: analisi statistica e probabilistica per una valutazione preliminare della pericolosità. In: *I convegno nazionale AIGA* (pp. 427–448)

Frattini P, Crosta GB, Sosio R (2009) Approaches for defining thresholds and return periods for rainfall-triggered shallow landslides. *Hydrological Processes: An International Journal* 23(10):1444–1460

Frattini P, Crosta GB, Carrara A (2010) Techniques for evaluating the performance of landslide susceptibility models. *Eng Geol* 111(1–4):62–72

Gabella M, Notarpietro R (2004) Improving operational measurement of precipitation using radar in mountainous terrain-part I: methods. *IEEE Geosci Remote Sens Lett* 1(2):78–83

Gariano SL, Brunetti MT, Iovine G, Melillo M, Peruccacci S, Terranova O, Guzzetti F (2015) Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy. *Geomorphology* 228:653–665

Glade T, Crozier M, Smith P (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

Godt JW, Baum RL, Chleborad AF (2006) Rainfall characteristics for shallow landsliding in Seattle, Washington, USA. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group* 31(1):97–110

Goodenough DJ (1974) The need for physical and psychophysical measures of system performance. In *Medical X-ray photo-optical systems evaluation* (Vol. 56, pp. 64–70). International Society for Optics and Photonics

Guzzetti F, Stark CP, Salvati P (2005) Evaluation of flood and landslide risk to the population of Italy. *Environ Manag* 36(1):15–36

Guzzetti F, Peruccacci S, Rossi M, Stark CP (2007) Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorol Atmos Phys* 98(3–4):239–267

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1):29–36

Hitschfeld W, Bordan J (1954) Errors inherent in the radar measurement of rainfall at attenuating wavelengths. *J Meteorol* 11(1):58–67

Hong M, Kim J, Jeong S (2018) Rainfall intensity-duration thresholds for landslide prediction in South Korea by considering the effects of antecedent rainfall. *Landslides* 15(3):523–534

Iadanza C, Trigila A, Napolitano F (2016) Identification and characterization of rainfall events responsible for triggering of debris flows and shallow landslides. *J Hydrol* 541:230–245

- Italian Ministry of Infrastructure and Transport (2012) Studio di valutazione dei costi sociali dell'incidentalità stradale. Ministero delle Infrastrutture e dei Trasporti, Direzione Generale per la sicurezza stradale. https://www.mit.gov.it/mit/mop_all.php?p_id=12919 (last access: 29/10/2020)
- Krajewski WF (1987) Cokriging radar-rainfall and rain gauge data. *J Geophys Res-Atmos* 92(D8):9571–9580
- Larsen MC, Simon A (1993) A rainfall intensity-duration threshold for landslides in a humid-tropical environment, Puerto Rico. *Geogr Ann* 75A(1-2):13–23
- Leonarduzzi E, Molnar P, McArdeell BW (2017) Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data. *Water Resour Res* 53(8):6612–6625
- Marino P, Peres DJ, Cancelliere A, Greco R, Bogaard TA (2020) Soil moisture information can improve shallow landslide forecasting using the hydrometeorological threshold approach. *Landslides* 17:2041–2054
- Marra F, Nikolopoulos EI, Creutin JD, Borga M (2014) Radar rainfall estimation for the identification of debris-flow occurrence thresholds. *J Hydrol* 519:1607–1619
- Martelloni G, Segoni S, Fanti R, Catani F (2012) Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides* 9(4):485–495
- Mathew J, Babu DG, Kundu S, Kumar KV, Pant CC (2014) Integrating intensity-duration-based rainfall threshold and antecedent rainfall-based probability estimate towards generating early warning for rainfall-induced landslides in parts of the Garhwal Himalaya, India. *Landslides* 11(4):575–588
- McRoberts EC, Morgenstern NR (1974) Stability of slopes in frozen soil, Mackenzie Valley, NWT. *Can Geotech J* 11(4):554–573
- Mirus BB, Becker RE, Baum RL, Smith JB (2018) Integrating real-time subsurface hydrologic monitoring with empirical rainfall thresholds to improve landslide early warning. *Landslides* 15(10):1909–1919
- Montgomery DR, Dietrich WE (1994) A physically based model for the topographic control on shallow landsliding. *Water Resour Res* 30(4):1153–1171
- Nikolopoulos EI, Borga M, Creutin JD, Marra F (2015) Estimation of debris flow triggering rainfall: Influence of rain gauge density and interpolation methods. *Geomorphology* 243:40–50
- Ochoa-Rodriguez S, Wang LP, Willems P, Onof C (2019) A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. *Water Resour Res* 55(8):6356–6391
- Piciullo L, Gariano SL, Melillo M, Brunetti MT, Peruccacci S, Guzzetti F, Calvello M (2017) Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. *Landslides* 14(3):995–1008
- Pierson TC (1980) Piezometric response to rainstorms in forested hillslope drainage depressions. *J Hydrol N Z* 19:1–10
- Postance B, Hillier J, Dijkstra T, Dixon N (2018) Comparing threshold definition techniques for rainfall-induced landslides: a national assessment using radar rainfall. *Earth Surf Process Landf* 43(2):553–560
- Provost F, Fawcett T (2001) Robust classification for imprecise environments. *Mach Learn* 42(3):203–231
- Schleiss M, Olsson J, Berg P, Niemi T, Kokkonen T, Thorndahl S, Nielsen R, Nielsen JE, Bozhinova D, Pulkkinen S (2020) The accuracy of weather radar in heavy rain: a comparative study for Denmark, the Netherlands, Finland and Sweden. *Hydrol Earth Syst Sci* 24(6):3157–3188
- Schmidt KM, Roering JJ, Stock JD, Dietrich WE, Montgomery DR, Schaub T (2001) The variability of root cohesion as an influence on shallow landslide susceptibility in the Oregon Coast Range. *Can Geotech J* 38(5):995–1024
- Seo DJ, Breidenbach JP, Johnson ER (1999) Real-time estimation of mean field bias in radar rainfall data. *J Hydrol* 223(3-4):131–147
- Simmons KM, Sutter D (2009) False alarms, tornado warnings, and tornado casualties. *Weather, Climate, and Society* 1(1):38–53
- Smith JA, Krajewski WF (1991) Estimation of the mean field bias of radar rainfall estimates. *J Appl Meteorol* 30(4):397–412
- Sorensen JH (2000) Hazard warning systems: review of 20 years of progress. *Natural hazards review* 1(2):119–125
- Staley DM, Kean JW, Cannon SH, Schmidt KM, Laber JL (2013) Objective definition of rainfall intensity-duration thresholds for the initiation of post-fire debris flows in southern California. *Landslides* 10(5):547–562
- Stevenson PC (1977) An empirical method for the evaluation of relative landslide risk. *Bull Int Assoc Eng Geol* 16:69–72
- Trigila A, Iadanza C, Spizzichino D (2010) Quality assessment of the Italian Landslide Inventory using GIS processing. *Landslides* 7(4):455–470
- Vaz T, Zêzere JL, Pereira S, Oliveira SC, Garcia RA, Quaresma I (2018) Regional rainfall thresholds for landslide occurrence using a centenary database. *Nat Hazards Earth Syst Sci* 18(4):1037–1054
- Wieczorek GF, Morgan BA, Campbell RH (2000) Debris-flow hazards in the Blue Ridge of central Virginia. *Environ Eng Geosci* 6(1):3–23
- Wilson JW, Brandes EA (1979) Radar measurement of rainfall—a summary. *Bull Am Meteorol Soc* 60(9):1048–1060
- Wu TH, McKinnell WP III, Swanston DN (1979) Strength of tree roots and landslides on Prince of Wales Island, Alaska. *Can Geotech J* 16(1):19–33

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10346-021-01707-4>.

G. Sala (✉) · **C. Lanfranconi** · **P. Frattini** · **G. B. Crosta**

Department of Earth and Environmental Sciences,
University Milano-Bicocca,
Milano, Italy
Email: g.sala39@campus.unimib.it

G. Rusconi

EG4 Risk SRL,
Milano, Italy