A Probabilistic Approach for Identifying Correlations between Landslides and Rainfall at Regional Scale

Michele Calvello[1] and Gaetano Pecoraro[2]

[1]Department of Civil Engineering, University of Salerno, Via Giovanni Paolo II, 132, Italy.
E-mail: mcalvello@unisa.it
[2]Department of Civil Engineering, University of Salerno, Via Giovanni Paolo II, 132, Italy.
E-mail: gpecoraro@unisa.it

Abstract: This study presents a methodology for the definition of probabilistic rainfall thresholds for landslide events at regional scale. The probabilistic analysis is based on the Bayesian theory and allows to consider the uncertainty of the data as well as to quantitatively estimate the reliability of the results. Data on landslide occurrences have been derived from the "FraneItalia" catalogue, an inventory of landslides retrieved from online Italian news. Correlations between landslides and rainfall are assessed by analyzing the events from March 2014 to December 2015 in the Emilia-Romagna region (Italy). The adopted territorial units are the eight warning zones of the region. Precipitation data have been gathered from the satellite-based NASA Global Precipitation Measurement (GPM) database, which contains gridded precipitation and precipitation-error estimates, with a half-hour temporal resolution and a 0.10-degree spatial resolution. The performed probabilistic analysis allows to highlight critical levels of rainfall corresponding to different probabilities, demonstrating the potential of the proposed probabilistic method to define objective and reproducible rainfall thresholds for landslide early warning purposes.

Keywords: Rainfall-induced landslides; early warning; probabilistic threshold; landslide catalogue; Italy.

1 Introduction

Rainfall-induced landslides are widespread and destructive natural phenomena occurring all around the world that often cause severe human and economic losses (Petley 2012). Landslide early warning systems (LEWS) are being increasingly applied as non-structural risk mitigation measures. Depending on the scale of their design and operation, they can be distinguished into: "local" systems and "territorial" systems (Pecoraro et al. 2019; Piciullo et al. 2018). Territorial LEWS typically deal with rainfall-induced landslides over appropriately defined warning zones. In these cases, warning models mainly rely on the monitoring and forecasting of meteorological variables and on empirical rainfall thresholds, which are defined by analyzing past rainfall events that have resulted or not in slope failures (Segoni et al. 2018). Developing empirical thresholds may be challenging for several reasons, such as: absence of a direct cause-effect relationship between rainfall and landslide initiation, homogeneity and completeness of available landslide catalogues, spatial and temporal resolution of rainfall data (Robbins 2016; Rossi et al. 2017).

In this study, a conceptual framework for the definition of probabilistic rainfall thresholds for landslides at regional scale has been developed. The main steps of the proposed approach are: collection of input data, correlation between landslides and rainfall events, and probabilistic analysis of the time series. The proposed procedure has been tested by analyzing the reported landslides from March 2014 to December 2015 within the eight warning zones defined for hydrogeological risk management in the Emilia-Romagna region (Italy).

This methodology should be seen as an innovative approach for the definition of rainfall thresholds for rainfall-induced landslides using open-access data from a non-conventional landslide inventory and rainfall satellite monitoring.

2 Materials and Methods

2.1 Study area and datasets

The study area is the Emilia-Romagna region in northern Italy (Figure 1). The northern and eastern portions of its territory are dominated by a wide flat area constituted by the alluvial plain of river Po, the longest Italian river. On the contrary, the southern and western portions of the region are occupied by the Apennines Chain, with a maximum altitude of 2165 m a.s.l. (Martelloni et al. 2012). The mountainous part of the Emilia-Romagna region is strongly affected by landslides, as more than 20% of the mountain territory is covered by active or dormant landslide deposits. The most frequent phenomena are rotational–translational slides, slow earth flows, and complex movements. However, the frequency of rapid shallow landslides has been markedly increasing in the last few years (Berti et al. 2012). This could be connected to the recent climatic trends in the Mediterranean area, which are characterized by shortest and more intense rainfalls, typically the main triggering factor of shallow

Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR)
Editors: Jianye Ching, Dian-Qing Li and Jie Zhang

Published by Research Publishing, Singapore.
ISBN: 978-981-11-2725-0; doi:10.3850/978-981-11-2725-0_IS13-1-cd

landslides and debris flows in the region (Floris et al. 2010). The operative warning system for flood and landslide risk is currently based on the division of the region in eight zones, defined following homogeneity criteria, including: physiography, lithology, precipitation regime, and administrative boundaries. The dataset used to analyze the case study includes information on landslide occurrences and rainfall measurements from March 2014 to December 2015.

Data on landslides come from "FraneItalia" (https://data.mendeley.com/datasets/zygb8jygrw/1), a geo-referenced openly available catalogue of Italian landslides retrieved from online news from 2010 onwards (Calvello and Pecoraro 2018). Landslide events are classified considering two numerosity categories: single landslide events (SLE), for records only reporting one landslide; and areal landslide events (ALE), for records referring to multiple landslides triggered by the same cause in the same geographic area. Each record of the catalogue is characterized by 40 unique fields, not all compulsory, which are grouped in 9 thematic tables: main info; spatial information; temporal information; landslide characteristics; consequences to people, structures, infrastructures, cars and other elements; and source of information. The database currently contains eight years of data (2010-2017) and it includes 5169 landslide events, most of them triggered by rainfall. Emilia-Romagna was affected by 102 rainfall-induced landslides in the period of analysis: 78 SLEs and 24 ALEs (Figure 1).

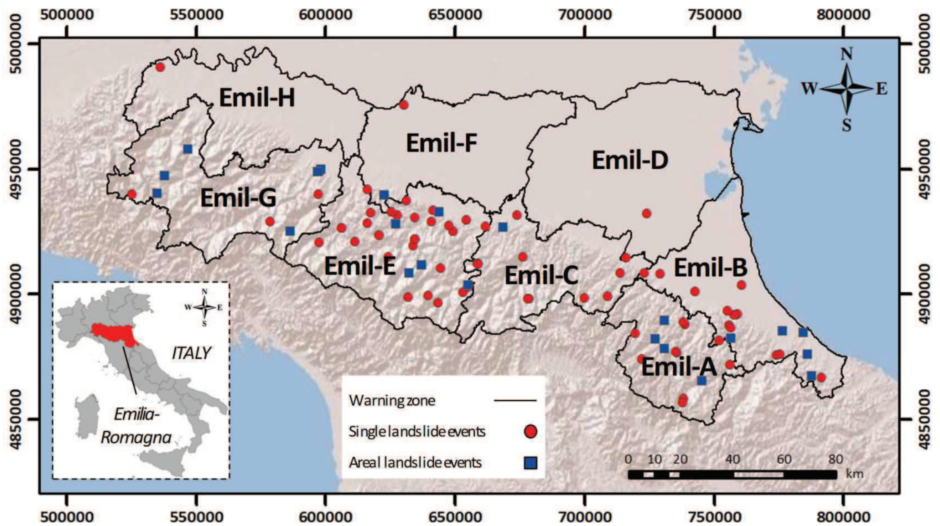

Figure 1. Shaded relief map of the eight warning zones of the Emilia-Romagna region showing the 102 rainfall-induced "FraneItalia" landslide records from March 2014 to December 2015, differentiated in single (red circles) and areal landslide events (blue squares). The inset shows the location of the Emilia-Romagna region in Italy.

The rainfall measurements were derived from the satellite-based Global Precipitation Measurement (GPM) mission, co-led by the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA) for weather and climate research purposes (Huffman et al. 2018). The mission was launched on 27 February 2014 and was a successor of the Tropical Rainfall Measuring Mission (TRMM), which provided data on heavy to moderate rainfall in Earth's tropics and subtropics from 1997 to 2015. GPM provides a wide variety of products retrieved combining data from active and passive instruments in the Integrated Multisatellite Retrievals for GPM (IMERG). Precipitation data used in this research have been derived from the IMERG version 5 (v05b), which includes gridded precipitation data collected every 30 min at a 0.1° × 0.1° (~10km × 10km) spatial resolution, currently covering the latitude band 60°N–60°S.

Satellite rainfall data retrieved from GPM database have been analyzed using Google Earth Engine (https://earthengine.google.com), a cloud-based platform for planetary-scale environmental data analysis allowing users to download and upload global satellite imagery as well as to carry out analysis on large datasets. For the purposes of this study, precipitation measurements have been aggregated at hourly temporal resolution and the mean rainfall values over each territorial unit have been calculated.

2.2 Two-dimensional Bayesian analysis

The definition of the probabilistic thresholds is based on the Bayesian theory, wherein the conditional probability can be defined as the probability of an event (herein, a landslide event) given that (by assumption, presumption, assertion or evidence) another event has occurred (herein, a rainfall event characterized by a certain magnitude, expressed in terms of rainfall parameters).

A Bayesian framework is well suited to handle multidimensional analysis with numerous variables, for example the combined effect of rainfall duration, rainfall intensity and antecedent precipitation on landslide triggering. However, multidimensional data are difficult to visualize and analyze, therefore researchers and practitioners often find it more efficient to restrict the analysis to two-dimensional cases (Berti et al. 2012).
In this study, the posterior landslide probability is evaluated considering the joint probability of the duration *(D)* and cumulated rainfall *(E)*, as follows:

$$P(L|D,E) = \frac{P(L)\times P(D,E|L)}{P(D,E)} \tag{1}$$

where *P(L|D,E)* is the posterior landslide probability, i.e. the conditional probability of a landslide event L given the joint probability of the two rainfall parameters *D* and *E; P(L)* is the prior probability, i.e. the probability of a landslide event *L; P(D,E|L)* is the likelihood, i.e. the conditional probability of *D* and *E* given the occurrence of a landslide event *L; P(D,E)* is the marginal probability, i.e. the joint probability of *D* and *E.* The needed probabilities are herein based on relative frequencies, which can be computed by:

$$P(L) = \frac{N_L}{N_R} \tag{2}$$

$$P(D,E) = \frac{N_{(D,E)}}{N_R} \tag{3}$$

$$P(D,E|L) = \frac{N_{(D,E|L)}}{N_L} \tag{4}$$

where N_L is the total number of landslide events that occurred in the period of analysis; N_R is the total number of rainfall events recorded in the period of analysis; $N_{(D,E)}$ is the number of rainfall events characterized by specific values of *D* and *E;* $N_{(A,B|L)}$ is the number of rainfall events characterized by specific values of *D* and *E* that resulted in landslides.

Following this approach, the frequency distributions for specific classes of rainfall events that resulted in landslides and did not result in landslides can be converted to posterior landslide probabilities. Any pair of parameters can be considered in two-dimensional Bayesian analysis (e.g., peak rainfall intensity, total event rainfall, antecedent rainfall), and their significance can be assessed by comparing the computed posterior landslide probability with the prior landslide probability.

Figure 2 shows an application of Eq. (1) to a sample dataset, considering duration *(D)* and cumulated rainfall *(E)* as rainfall parameters. All the thirty rainfall events are plotted in the duration-cumulated rainfall plane, which is divided into four regions delimited by the *D* and *E* values (Figure 2a). Eq. (1) is then computed separately for each region obtaining probabilistic information in the *DE* plane (Figure 2b). For example, in the upper-left cell 3 rainfall events out of 6 resulted in landslides, which means that P*(D,E|L) = 3/10 = 0.30* and *P(D,E) = 6/30 = 0.20.* The prior landslide probability is *P(L) = 0.33* and the posterior landslide probability is *P(L|D,E) = 3/10 = 0.50* (Figure 5.8b).

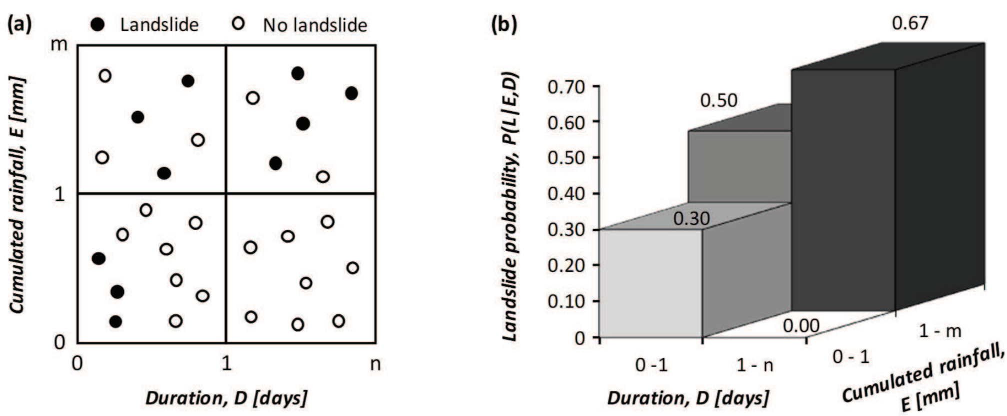

Figure 2. Example of two-dimensional Bayesian analysis. (a) Rainfall intensity-duration plot showing rainfall that did and did not result in landslides. (b) Histogram of conditional landslide probability for four different combinations of duration and cumulated rainfall.

3 Results and Discussion

3.1 Correlation between landslides and rainfall events

The definition of the correlation between landslides and rainfall events in Emilia-Romagna required the reconstruction of the rainfall events in order to pass from a series of hyetographs to a point cloud in a graph reporting triggering and non-triggering combinations of rainfall parameters. To this aim, an algorithm widely used in the literature (Segoni et al. 2018 and references therein) has been adopted. The procedure needs three main steps to define a rainfall event based on its attributes (i.e., D and E). In the first step, all the isolated rainfall events are detected considering a minimum period without rain. To account for different meteorological regimes in Italy, two minimum periods are considered: a two-day (48 h) interval for the "warm" spring-summer period (April-September) and a four-day (96 h) interval for the "cold" period (October-March). Successively, isolated rainfall measurements that do not exceed a minimum value E_R=1 mm are excluded, because they are considered not relevant for possible landslide initiation. Finally, the triggering and non-triggering rainfall conditions are identified. Regarding triggering rainfall conditions, landslides typically occur before the end of the rainfall event. In these cases, the rainfall after the landslide occurrence cannot be considered relevant for the initiation of the slope failure. On the contrary, when landslides fail after the end of the rainfall period, the rainfall associated to the landslide corresponds to the cumulated rainfall of the entire event.

Following this approach, 1029 rainfall conditions *(D,E)* have been reconstructed and plotted in log-log coordinates (Figure 3). The 102 rainfall conditions responsible for the triggering of the 78 SLEs (red circles in Figure 3) and the 24 ALEs (blue squares in Figure 3) are in the range of duration $1 \leq D \leq 67\ h$ and in the range of cumulated rainfall $1.1 \leq E \leq 969\ mm$. Of course, the non-triggering rainfall conditions reconstructed in the same period are 927 (grey circles in Figure 3). These rainfall events are in the ranges of $1 \leq D \leq 79.5\ h$ and $1 \leq E \leq 963.5\ mm$. The graph does not show a clear distinction between triggering and non-triggering rainfall events, thus the application of conventional methods for the definition of a rainfall threshold can be extremely difficult and a probabilistic approach seems to be more appropriate. It is worth mentioning that the probabilistic analyses have been herein performed grouping SLEs and ALEs in a unique dataset (ALEs affecting several territorial units have been preliminarily separated and distributed among the territorial units where they occurred).

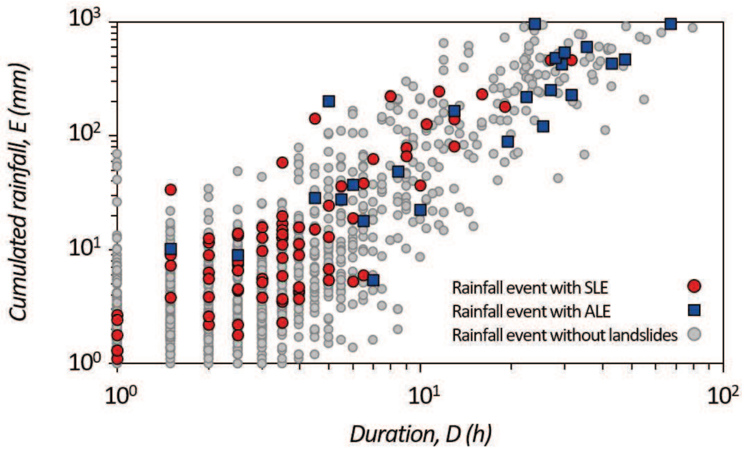

Figure 3. Rainfall duration (D) vs. cumulated rainfall (E) in Emilia-Romagna from March 2014 to December 2015. Red circles are 78 ED rainfall conditions associated with the triggering of SLEs. Blue squares are the 24 ED rainfall conditions associated with the triggering of ALEs. Grey circles are 927 rainfall events representative of non-triggering rainfall conditions. Data are in log-log coordinates.

3.2 Definition of the probabilistic threshold

The definition of the probabilistic threshold is based on a two-dimensional Bayesian analysis evaluating the conditional probability of landslide occurrence given the joint probability of *D* and *E*. Following the procedure described in Section 2.2, Eq. (2) has been applied in order to compute the prior landslide probability, *P(L)*, using the data reported in Section 3.1:

$$P(L) = \frac{N_L}{N_R} = \frac{102}{1029} = 0.10 \tag{5}$$

The *ED* space reported in Figure 3 has been divided in 6x10 cells and the posterior landslide probabilities, *P(L|D,E)*, have been calculated by applying Eq. (1). The results have been interpolated onto a 2D plot so that the probabilities across the log-log plane can be better visualized. Figure 4 displays that long-duration, high-accumulation rainfall events show the highest landslide probabilities (*P(L|D,E)*=0.5). Besides, three secondary peaks (*P(L|D,E)*>0.2) can be also observed: two of them in the proximity of the main peak and another in correspondence of short-duration, mid-accumulation rainfall events. This could suggest that two different types of rainfall events are more likely to trigger landslides.

Landslide probabilities have been further processed in order to draw lines of roughly equal landslide probability (i.e., isolines) in the log-log plane. They represent rainfall events characterized by different duration and magnitude, which result in the same probability of landslide occurrence and can be used as thresholds for early warning purposes. To this aim, Berti et al. (2012) stated that a reasonable criterion for setting the threshold would be the observation of an abrupt increase in the probability of failure, which indicates a radical change in the state of the system. In this case, $P(L|D,E) = 0.15$ can be considered as an appropriate threshold, because above this line the landslide probability rapidly increases for different ranges of duration and cumulated rainfall.

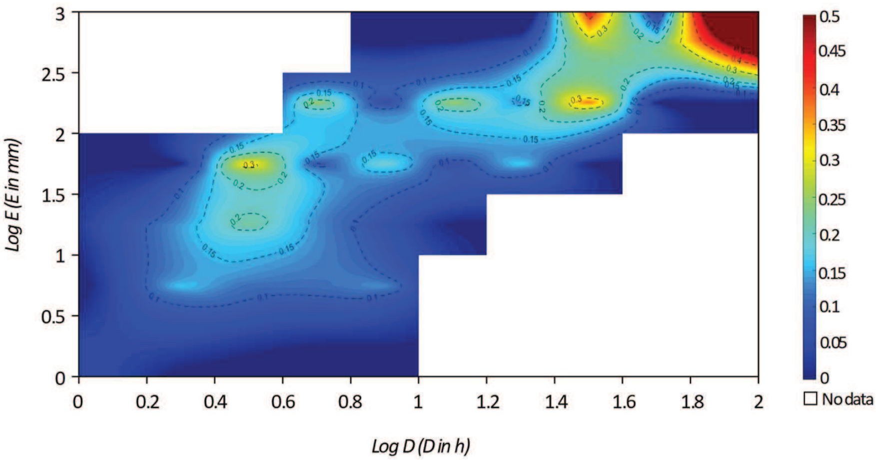

Figure 4. Posterior landslide probabilities obtained considering the rainfall events reported in Figure 2. Lines of equal probability are also drawn.

3.3 Comparison with other rainfall thresholds

For validation purposes, the new probabilistic rainfall threshold obtained has been quantitatively compared with other *ED* rainfall thresholds defined for the same region: the environmental thresholds defined by Peruccacci et al. (2017) and the regional threshold derived by the probabilistic model developed by Berti et al. (2012). It should be stressed that the direct comparison of these regional thresholds is not appropriate, as they were defined using different methods and techniques, different landslide and rainfall information, and adopting different criteria to identify rainfall conditions that resulted in landslides. Considering these limitations, a back analysis has been carried out using the same rainfall and landslide datasets employed for the determination of the probabilistic threshold in order to perform a quantitative comparison of the three approaches. To this aim, thresholds have been considered as binary classifiers of rainfall conditions that may result (or not result) in landslides. Adopting this assumption, a set of contingencies scores—i.e., true positives *(TP),* true negatives *(TN),* false positives *(FP),* and false negatives *(FN)*—derived by standard contingency tables has been considered (Wilks 1995).

Table 1 summarizes the four contingencies scores for the different thresholds defined for Emilia-Romagna (only the lower and upper envelopes of the thresholds defined by Peruccacci et al. (2017) have been considered). The probabilistic threshold ($T_{15,P}$) allows maximizing the number of *TP* (78), concurrently not increasing excessively the number of *FP* (217) with respect to the other thresholds. Although $T_{5,UE}$ and $T_{5,R}$ show the lowest values of *FP* (203 and 212, respectively), they miss a relevant number of landslide events (65 and 64, respectively). On the contrary, $T_{5,LE}$ results in a relevant number of *FP* (about 50% higher than the probabilistic threshold).

The overall best performance of $T_{15,P}$ can be explained taking into account that it is more flexible than conventional thresholds typically represented by linear classifiers. This feature could be extremely useful in areas where slope response to rainfall is quite complex, i.e., when landslides are triggered by different rainfall conditions.

Table 1. Contingencies scores calculated for the probabilistic threshold *($T_{15,P}$)* and for other thresholds proposed in the literature (calculated at 5% exceedance probability) for Emilia-Romagna. Best values are shown in italics.

Threshold	Label	TP	FN	FP	TN
Probabilistic threshold	$T_{15,P}$	*78*	*24*	217	710
Environmental thresholds – LE *(Peruccacci et al. 2017)*	$T_{5,LE}$	58	44	320	607
Environmental thresholds – UE *(Peruccacci et al. 2017)*	$T_{5,UE}$	37	65	*203*	*724*
Regional threshold *(Berti et al. 2012)*	$T_{5,R}$	38	64	212	715

4 Conclusions

In this study a Bayesian approach has been developed for the definition of a probabilistic rainfall threshold using a landslide inventory retrieved from online news and satellite-based rainfall measurements. The analyses were conducted within the eight weather warning zones of the Emilia-Romagna region (Italy) from March 2014 to December 2015. The results indicate that two rainfall conditions are more likely to trigger landslides: long-duration high-accumulation rainfall events, and short-duration mid-accumulation rainfall events. This is probably due to the heterogeneity of the calibration dataset. Indeed, although all the records refer to shallow rainfall-induced landslides, they significantly differ in terms of triggering mechanism (different rainfall conditions lead to landslide initiation) and magnitude (both single and areal events have been grouped in the dataset).

In addition, a back analysis has been performed to assess the performance of the probabilistic threshold defined herein (posterior landslide probability equal to 0.15) with other rainfall thresholds proposed in the literature for the Emilia-Romagna region. The quantitative comparison revealed an overall good performance of the probabilistic threshold in predicting landslide events triggered by significantly different rainfall conditions. Although the performance of the probabilistic model can be further refined considering longer datasets and employing more than one threshold, the preliminary results achieved herein clearly allow to highlight the potential of the proposed model for landslide early warning purposes.

Acknowledgments

This work was financially supported by the PhD programme of the Civil Engineering Department of the University of Salerno, for GP.

References

Berti, M., Martina, M.L., Franceschini, S., Pignone, S., Simoni, A., and Pizziolo, M. (2012). Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. *Journal of Geophysical Research: Earth Surface*, 117, F04006.

Calvello, M. and Pecoraro, G. (2018). FraneItalia: a catalog of recent Italian landslides. *Geoenvironmental Disasters*, 5(13).

Floris, M., D'Alpaos, A., Squarzoni, C., Genevois, R., and Marani, M. (2010). Recent changes in rainfall characteristics and their influence on thresholds for debris flow triggering in the Dolomitic area of Cortina d'Ampezzo, north-eastern Italian Alps. *Natural Hazards and Earth System Sciences*, 10, 571–580.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd C, Nelkin, E.J., Sorooshian, S., Tan, J., and Xie P. (2018). *NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD), version 5.2*, NASA, 30 .
Available at: https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V5.2_0.pdf (accessed 14 December 2018).

Martelloni, G., Segoni, S., Fanti, R., and Catani F. (2012). Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides*, 9, 485–495.

Pecoraro, G., Calvello, M., and Piciullo, L. (2019). Monitoring strategies for local landslide early warning systems. *Landslides*, 16(2), 213-231.

Peruccacci, S., Brunetti, M.T., Gariano, S.L., Melillo, M., Rossi, M., and Guzzetti F. (2017). Rainfall thresholds for possible landslide occurrence in Italy. *Geomorphology*, 290, 39–57.

Petley, D. (2012). Global patterns of loss of life from landslides. *Geology,* 40(10), 927–930.

Piciullo, L., Calvello, M., and Cepeda, J.M. (2018). Territorial early warning systems for rainfall-induced landslides. *Earth-Science Reviews*, 179, 228–247.

Robbins, J.C. (2016). A probabilistic approach for assessing landslide-triggering event rainfall in Papua New Guinea, using TRMM satellite precipitation estimates. *Journal of Hydrology*, 541, 296–309.

Rossi, M., Luciani, S., Valigi, D., Kirschbaum, D., Brunetti, M.T., Peruccacci, S., and Guzzetti F. (2017). Statistical approaches for the definition of landslide rainfall thresholds and their uncertainty using rain gauge and satellite data. *Geomorphology*, *285,16–27.*

Segoni, S., Piciullo, L., and Gariano, S.L. (2018). A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides*, 15, 1483–1501.

Wilks, D.S. (1995). *Statistical Methods in the Atmospheric Sciences*. Academic Press, USA, 467 pp.