Environmental Modelling & Software 105 (2018) 230–243

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A tool for the automatic calculation of rainfall thresholds for landslide occurrence

Massimo Melillo*, Maria Teresa Brunetti, Silvia Peruccacci, Stefano Luigi Gariano, Anna Roccati, Fausto Guzzetti

Istituto di Ricerca per La Protezione Idrogeologica, Consiglio Nazionale Delle Ricerche, Via Madonna Alta 126, 06128 Perugia, Italy

ARTICLE INFO

Article history:
Received 14 December 2017
Received in revised form
23 March 2018
Accepted 30 March 2018

Keywords:
Rainfall event
Rainfall modelling
Failure
Evapotranspiration
Liguria

ABSTRACT

Empirical rainfall thresholds are commonly used to forecast landslide occurrence in wide areas. Thresholds are affected by several uncertainties related to the rainfall and the landslide information accuracy, the reconstruction of the rainfall responsible for the failure, and the method to calculate the thresholds. This limits the use of the thresholds in landslide early warning systems. To face the problem, we developed a comprehensive tool, CTRL–T (Calculation of Thresholds for Rainfall-induced Landslides–Tool) that automatically and objectively reconstructs rainfall events and the triggering conditions responsible for the failure, and calculates rainfall thresholds at different exceedance probabilities. CTRL–T uses a set of adjustable parameters to account for different morphological and climatic settings. We tested CTRL–T in Liguria region (Italy), which is highly prone to landslides. We expect CTRL–T has an impact on the definition of rainfall thresholds in Italy, and elsewhere, and on the reduction of the risk posed by rainfall-induced landslides.

Software availability

Name: CTRL–T (Calculation of Thresholds for Rainfall-induced Landslides–Tool)
Developer: Massimo Melillo
Contact Address: CNR-IRPI, via Madonna Alta 126, 06128 Perugia, Italy
Email: massimo.melillo@irpi.cnr.it
Program language: R
Software requirement: R
Integrated Development Environment (IDE) suggest: RStudio
Source code: http://geomorphology.irpi.cnr.it/tools/rainfall-events-and-landslides-thresholds/ctrl-algorithm/ctrl-code/CTRL_code.R/

1. Introduction

Landslides are natural and human-induced phenomena that affect all continents, playing an important role in the evolution of landscapes and posing a serious threat to the population (Keefer and Larsen, 2007; Nadim et al., 2006, 2013; Petley, 2012). Rainfall is a recognized trigger of landslides, and this explains why there is a vast scientific literature on the relationship between rainfall and landslide occurrence. At regional and global scales, empirical rainfall thresholds are among the most used tools for the prediction of rainfall-induced slope failures. Several authors have proposed different methods for the calculation of rainfall thresholds through the statistical analysis of empirical distributions of rainfall conditions that have presumably resulted in landslides – including cumulated event rainfall vs. rainfall duration or mean rainfall intensity vs. rainfall duration (e.g., Aleotti, 2004; Guzzetti et al., 2007, 2008; Brunetti et al., 2010; Berti et al., 2012; Giannecchini et al., 2012; Martelloni et al., 2012; Peruccacci et al., 2012; Staley et al., 2013; Segoni et al., 2014; Rosi et al., 2016; Galanti et al., 2017). Some authors have considered both rainfall conditions that have and have not resulted in landslides and have used optimization techniques to define the optimal thresholds (e.g., Berti et al., 2012; Staley et al., 2013; Peres and Cancelliere, 2014). On the other hand, attempts to predict the occurrence of rainfall-induced landslides by means of a physically-based approach are present in the scientific literature (e.g., Lepore et al., 2017; Alvioli et al., 2014; Alvioli and Baum, 2016; An et al., 2016; Peres and Cancelliere, 2016).

Empirical rainfall thresholds are affected by several uncertainties, including: (i) the availability and quality of the rainfall

* Corresponding author.
E-mail address: massimo.melillo@irpi.cnr.it (M. Melillo).

https://doi.org/10.1016/j.envsoft.2018.03.024

measurements and of the landslide information (Guzzetti et al., 2007; Berti et al., 2012; Peruccacci et al., 2012; Nikolopoulos et al., 2014; Gariano et al., 2015; Marra et al., 2017; Peres et al., 2017); (ii) the characterization of the rainfall event responsible for the landslides (Guzzetti et al., 2007; Iadanza et al., 2016); (iii) the heuristic or statistical methods used to determine the thresholds (Peruccacci et al., 2012, 2017; Vennari et al., 2014). As for the first point, uncertainty exists especially in rainfall series containing large data gaps. Depending on the information source, uncertainty can also affect the geographical and the temporal location of the failure. Gariano et al., 2015 found that even a small (1%) lack of landslide information can result in a significant decrease in the performance of a threshold-based prediction model. Uncertainty in the characterization of rainfall data has a strong impact on the identification of the thresholds, often resulting in their underestimation, leading to a high number of false alarms in early warning system applications (Nikolopoulos et al., 2014; Marra et al., 2017; Peres et al., 2017). Regarding the second point, Melillo et al., 2015 highlighted how standards for measuring landslide-triggering rainfall conditions are still lacking or insufficiently formalized in literature. Indeed, how the rainfall responsible for the landslide occurrence is calculated and whether its definition is reliable it is rarely reported, thus reducing the possibility of comparing different thresholds. Concerning the last point, the majority of empirical rainfall thresholds available in the literature are calculated using subjective and scarcely repeatable methods. Only few attempts were recently made to conceive procedures for an objective and reproducible definition of thresholds (Brunetti et al., 2010; Martelloni et al., 2012; Staley et al., 2013; Segoni et al., 2014; Vessia et al., 2014; Melillo et al., 2015; Piciullo et al., 2017; Peruccacci et al., 2017). We maintain that the quantitative identification of the landslide-triggering rainfall and the definition of reliable thresholds are fundamental steps towards a well-founded landslide prediction.

In this work, we upgraded the algorithm proposed by Melillo et al. (2015, 2016), introducing new criteria to select automatically (i) the representative rain gauge (i.e., the most representative measuring station suitable to reconstruct the landslide-triggering rainfall) and (ii) the rainfall conditions responsible for landslides. The algorithm standardizes the actions performed by an investigator that defines rainfall events starting from series of rainfall records and landslide occurrence dates. In addition, the algorithm, modelling the cumulated event rainfall, identifies the rainfall conditions responsible for the observed failures and calculates rainfall thresholds at different exceedance probabilities (Brunetti et al., 2010; Peruccacci et al., 2012). The algorithm is implemented in a tool (CTRL–T, Calculation of Thresholds for Rainfall-induced Landslides–Tool) written in R open-source software (Appendix A).

The paper is organized as follows. First, we describe the tool and the main upgrades and improvements herein proposed in the algorithm (Section 2), then we describe data and study area (Section 3) and some specific parameters (Section 4). In Section 5, we show the results obtained from its application in the study area (Liguria, northwestern Italy). This is followed, in Section 6, by a discussion on the main advantages of the tool, and its potential application for the reconstruction and assessment of the rainfall conditions that can initiate landslides in wide geographical areas. We conclude (Section 7) summarizing the main findings of the work.

2. Description of CTRL–T

CTRL–T contains an improved version of the algorithm proposed by Melillo et al. (2015, 2016). The former version of the algorithm allowed: (i) the reconstruction of distinct rainfall events (by pre-processing and analyzing rainfall measurements and aggregating rainfall sub-events); (ii) the identification of multiple rainfall conditions responsible for the triggering of documented landslides; (iii) the definition of rainfall thresholds for possible landslide occurrence. The former algorithm used pre-defined parameters to account for different seasonal and climatic conditions. CTRL–T includes new criteria to reconstruct automatically the rainfall conditions responsible for landslides and to define rainfall thresholds at different exceedance probabilities. Fig. 1 illustrates the logical framework of the improved algorithm. The "INPUT" data

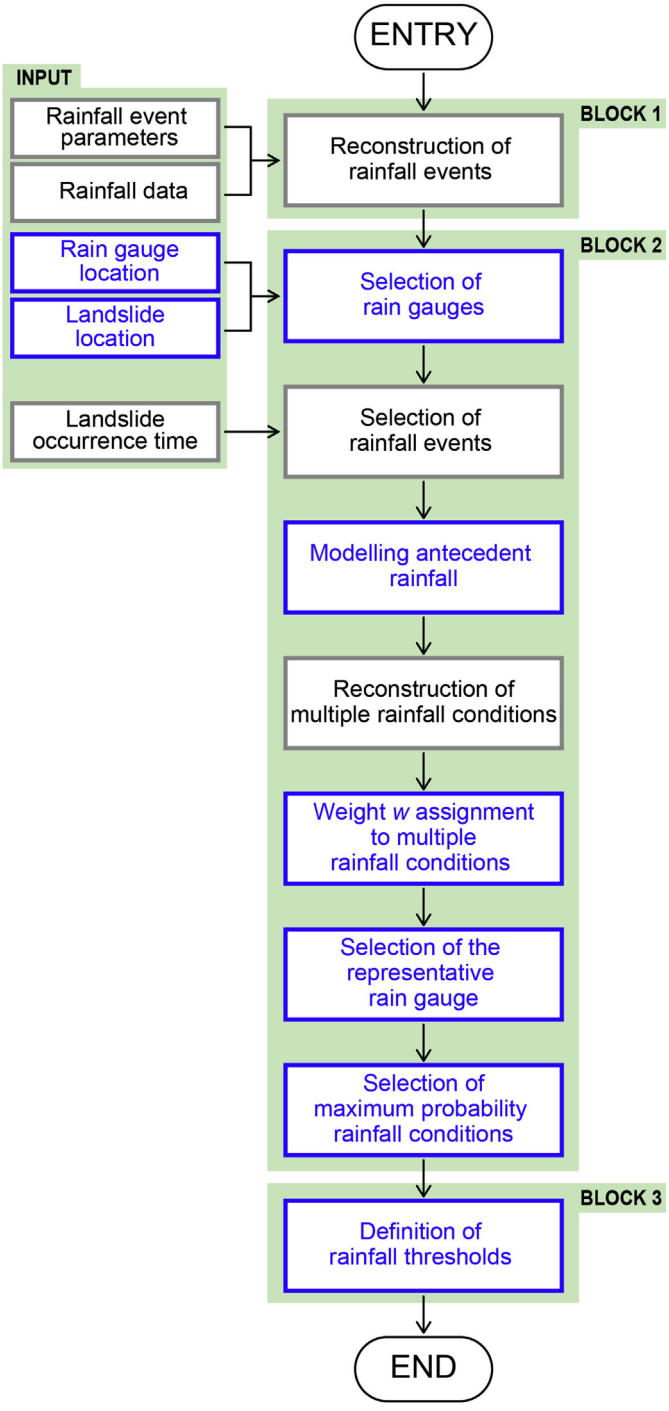

Fig. 1. Logical framework of the algorithm in CTRL–T. Grey and blue-bordered boxes represent actions already implemented in the previous release of the algorithm and the new additional actions, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

is composed by: (i) setting parameters of the rainfall events, (ii) rainfall data, (iii) rain gauge locations, (iv) landslide locations, and (v) landslide occurrence times. In Fig. 1, grey-bordered boxes represent actions already implemented in the previous version of the algorithm, while blue-bordered boxes highlighted the new additional actions.

The algorithm is divided into three main logical blocks (Fig. 1). The "BLOCK 1" executes the reconstruction of the rainfall events (*RE*) using the setting parameters and the rainfall series obtained by the rain gauges located in the study area. A rainfall event is a period of continuous rainfall, or an ensemble of periods of continuous rainfall, separated from the preceding and the successive events by dry (no-rain) periods. The parameters are selected according to the climate conditions of the area (Melillo et al., 2015; see Section 4).

Using the information on the location of rain gauges and landslides provided by the "INPUT" section, "BLOCK 2" picks out the rain gauges closest to each landslide. Details about the selection of rain gauges are reported in Section 2.1. In the following action of the same block, for each selected rain gauge, the algorithm identifies the rainfall event associated with the landslide. Adopting a modelling of the cumulated event rainfall described in Section 2.2, and using criteria reported in Melillo et al. (2015), the algorithm reconstructs the single or the multiple rainfall conditions (*MRC*) likely responsible for each failure. *MRC* can be a (D_L,E_L) pair of rainfall event duration (D_L) and cumulated event rainfall (E_L), or a set of two or more pairs. Through an empirical relation that includes the distance between the rain gauge and the landslide, D_L and E_L (see Section 2.3), a weight *w* is assigned to each pair of the *MRC* data set. For each landslide, among the selected rain gauges, the *MRC* corresponding to the maximum *w* is that likely responsible for the failure.

In the "BLOCK 3", the set of the *MRC* associated to all the available landslides is used to calculate the rainfall thresholds (Section 2.4).

In the following sections, we describe in detail the new actions included in the improved version of the algorithm.

2.1. Rain gauge selection procedure

In the previous release of the algorithm (Melillo et al., 2015), the selection of the representative rain gauge was done manually through multiple queries to the database of rainfall measurements. In the new release, the algorithm was improved to select automatically the representative rain gauge for a specific landslide from a pool of rain gauges close to the landslide (Fig. 2).

The geographical locations of the landslides and of the rainfall stations are a necessary input of CTRL–T (Fig. 1). For each failure, nearby rain gauges are located in a circular area (buffer) centered on the landslide location and with a parametrized radius (R_b, in Fig. 2). R_b depends chiefly on the morphological settings of the study area and on the rain gauge density. As an example, in mountain regions where the altitude changes abruptly, the buffer radius should not exceed 5 km, whereas in flat regions with low rain gauge density it can increase to 15 km. The representative rain gauge used to reconstruct the rainfall responsible of the failure is identified successively after an objective analysis of the reconstructed *MRC* conditions (Section 2.3).

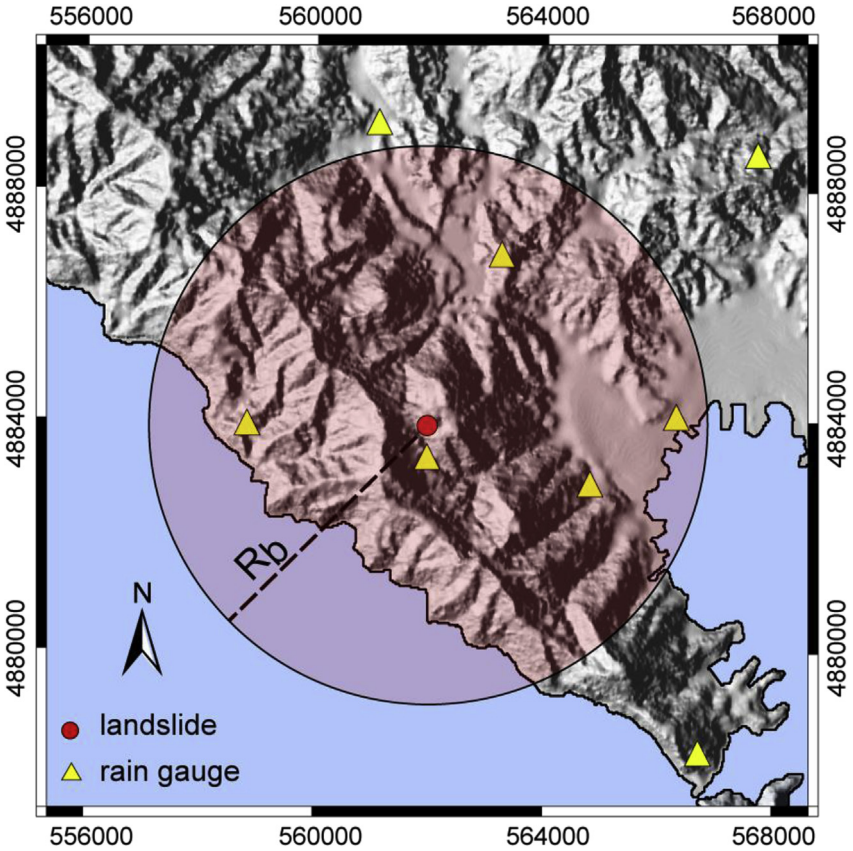

Fig. 2. Selection of rain gauges (yellow triangles) in a circular area (buffer) centered in the landslide location (red dot) with a parametrized radius (dashed black line). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

2.2. Modelling the cumulated rainfall responsible for the failure

In the previous release of the algorithm, the contribution of the rainfall to the landslide occurrence was fully included regardless of the time elapsed since the start of the rain; hence, any possible effect of the soil water saturation was neglected. To overcome this problem, CTRL–T models the rainfall by applying an arbitrary constant decay factor (e.g., $k = 0.84$) to the cumulated event rainfall E_L, as proposed by Crozier (1999):

$$E_L = E_L(0) + kE_L(1) + k^2E_L(2) + \ldots + k^NE_L(N) = \sum_{i=0}^{N} k^i[E_L(i)] \quad (1)$$

where $E_L(0)$ is the cumulated rainfall in the 24 h before the landslide occurrence time (t_L), $E_L(i)$ is the cumulated rainfall in the 24 h of the i-th day before t_L, and N is the duration of the rainfall event in days (i.e., steps of 24 h). As an example, the rainfall contribution to the amount of rainfall in the 5th day before the landslide reduces to about one half (49.8%) of the measured rain. The k decay factor can be changed according to the soil moisture drainage of the local study area.

2.3. Weight assignement

For each landslide, the algorithm proposed by Melillo et al. (2015) identified a variable number n of single ($n = 1$) or multiple rainfall conditions likely responsible for the failure. All the reconstructed (D_L,E_L) pairs were considered equally-probable (i.e., with the same weight $w = 1$) landslide triggering conditions. In a successive release of the algorithm (Melillo et al., 2016), the rainfall conditions were selected with a weight w inversely proportional to n ($w = 1/n$). With that assumption, the weight of a rainfall condition would not depend on the D_L and E_L variables, but only on its multiplicity n ($w = f(n)$). Here, we propose a new empirical meaning of w, which is proportional to the inverse square distance between the rain gauge and the landslide (d-2), the cumulated rainfall (EL), and the rainfall mean intensity (ELDL-1):

$$w = f(d, E_L, D_L) = d^{-2}E_L^2D_L^{-1} \quad (2)$$

The weight is attributed to each (D_L,E_L) pair of the *MRC* data set. For multiple (D_L,E_L) pairs (i.e., rainfall conditions with increasing duration and increasing cumulated rainfall), when the difference in the cumulated rainfall E_L between one pair and the subsequent is less than 10%, the weight attributed to the latter pair is null ($w = 0$) and this condition is removed from the *MRC* data set. Similarly, w is null for rainfall conditions having a delay between the rainfall ending time and the landslide occurrence time longer than an assigned time set to 48 h. This last requirement prevents the use of wrong information (i.e., incorrectly dated landslides) in the threshold definition. Indeed, we expect that the considered landslides are mostly shallow and therefore occur within a short delay after the triggering rainfall.

For each landslide, w is used to identify the representative rain gauge, considering both geographical and rainfall features, and to determine the probability of the single or multiple rainfall conditions to be adopted for the calculation of rainfall thresholds. For a pool of stations enclosed in the radius R_b, the representative rain gauge is the one for which the (D_L,E_L) pair has the highest w. In case of multiple pairs, each w is normalized to the sum of the individual weights, whereas is set to one in case of a single condition. CTRL–T calculates rainfall thresholds for *MRC* and *MPRC* (maximum probability rainfall conditions) data sets, where *MPRC* is the subset of the (D_L,E_L) pairs with the highest weights.

2.4. Definition of rainfall thresholds

To define empirical rainfall thresholds and their associated uncertainties, CTRL–T adopts a bootstrapping statistical technique (Peruccacci et al., 2012) and a frequentist method (Brunetti et al., 2010) by sampling the weighted rainfall conditions that have triggered landslides. The general form of the threshold curves is a power law:

$$E = (\alpha \pm \Delta\alpha) \times D^{(\gamma \pm \Delta\gamma)} \quad (3)$$

where E is the cumulated event rainfall (in mm), D is the rainfall event duration (in h), α is a scaling constant (the intercept), γ is the shape parameter (that defines the slope of the power law curve), and $\Delta\alpha$ and $\Delta\gamma$ represent the uncertainties of α and γ, respectively.

In Eq. (3), α and γ are the mean values of the parameters resulting from the calculation of thresholds of 5000 synthetic series generated by the algorithm. $\Delta\alpha$ and $\Delta\gamma$ are the standard deviation of α and γ, respectively. Each synthetic series contains the same number n of landslides as the original data set but selected randomly with replacement (bootstrap technique). To calculate the thresholds, a single (D_L,E_L) pair is associated to each landslide. For each landslide in the individual synthetic series, the algorithm samples randomly – with a probability w – a single rainfall condition from the *MRC* data set. The extracted (D_L,E_L) pairs of the n landslide are used to define the thresholds. The algorithm also uses the rainfall conditions with the maximum w to define thresholds for the *MPRC* data set.

The output of the bootstrapping technique consists of 5000 synthetic series of m (D_L,E_L) pairs. Analysis of the m synthetic series allows calculating the mean value and the uncertainty associated with the threshold parameters (α and γ) and their respective uncertainties ($\Delta\alpha$, $\Delta\gamma$).

3. Study area and data

We tested the tool using rainfall and landslide data available to us for the Liguria region (5410 km^2) in northwestern Italy. We used hourly rainfall measurements collected in the 15-year period from March 2001 to December 2014 by a network of 172 rain gauges operated by the *Osservatorio Meteo Idrologico della Regione Liguria* (OMIRL). Fig. 3 portrays the geographical distribution of the rain gauges (yellow triangles), having an average density of about one rain gauge every 31 km^2.

Moreover, we exploited geographical and temporal information on 561 rainfall-induced landslides occurred in Liguria between October 2004 and November 2014 (red dots in Fig. 3). Among those, there are 73 rock falls, 16 mudflows, 71 other types of landslides (e.g, earth flows, debris blows, etc.), and 401 unspecified shallow landslides. On-line chronicles and reports of local fire brigades were the exclusively information source for 337 (60%) and 51 (9%) landslides, respectively. The remaining 173 events (31%) were reconstructed using information from both sources. Generally, these last events have both a high geographical and temporal accuracy.

4. Computation of the regional parameters

In a previous version of the algorithm applied in the Sicily Region (southern Italy), Melillo et al. (2015) used a heuristic approach to separate two consecutive rainfall events. They selected a minimum dry (i.e., with no rain) interval of 2 days (48 h) in the “warm” period C_W (April–October), and a minimum of 4 days (96 h) in the “cold” period, C_C (November–March). In the same paper, the time

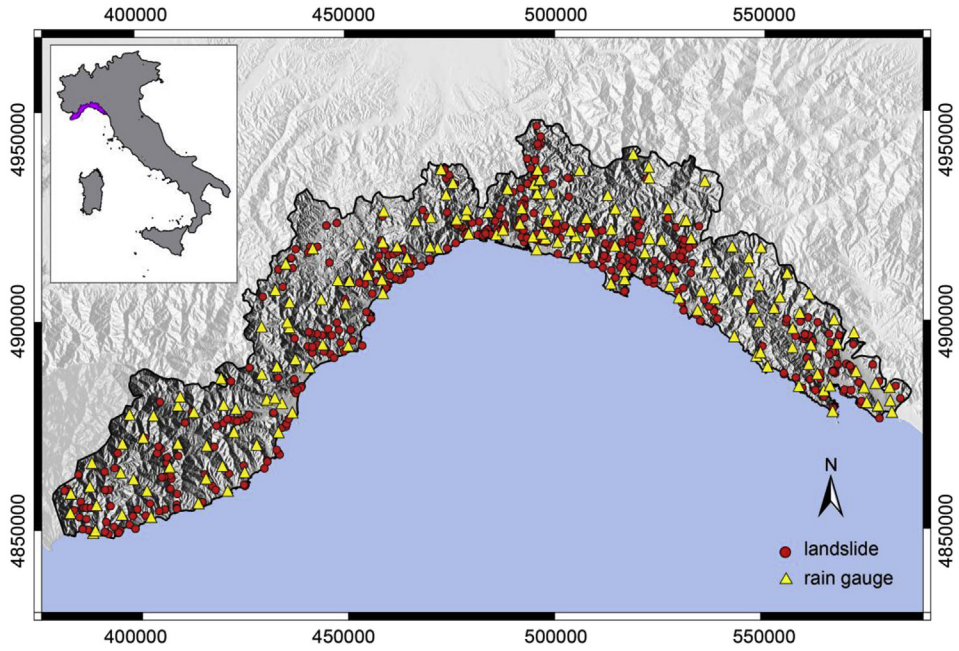

Fig. 3. Map showing the geographical location of 561 rainfall-induced landslides occurred in the Liguria region (northwestern Italy) between October 2004 and November 2014 (red dots). Yellow triangles show the location of 172 rain gauges. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

required for the soil to dry out was assumed inversely proportional to the amount of real evapotranspiration (*RET*), and they found a ratio$R = RET(C_W)/RET(C_C) = 2$. However, depending on the local seasonal and climatic conditions, the length (number of months) of the C_W and C_C periods may vary and may lead *R* to vary accordingly (Thorp, 1986). For this reason, among several methods proposed for evaluating the evapotranspiration (Alley, 1984; Yates, 1996; Naoum and Tsanis, 2003; Guo et al., 2016; Pumo et al., 2017), we calculated the *R* value adopting the monthly soil water balance (MSWB) model (Thornthwaite, 1948; Thornthwaite and Mather, 1957; Melillo, 2009; Di Matteo et al., 2011; Dragoni et al., 2015) and we determined the length (in months) of the C_W and C_C periods. The MSWB model is a simplified approach (with only few input requirements), and is commonly used to obtain reliable estimates of the water balance components (e.g., potential and real evapotranspiration). Input data of the MSWB model are: (i) the latitude; (ii) the average monthly rainfall, $\overline{E}_m$; (iii) the average monthly temperature, $\overline{T}_m$; and (iv) the maximum field capacity or soil water storage (SWS). We calculated $\overline{E}_m$using data available from the network of 172 rain gauges operated by the OMIRL in the Liguria Region (see section 3) and $\overline{T}_m$ from data provided by the website of *Istituto Superiore per la Protezione e la Ricerca Ambientale* (www.scia.isprambiente.it) in the period from January 2000 to December 2014. Fig. 4a shows a graph (Bagnouls-Gaussen diagram) with the average monthly rainfall (grey histogram) and the average monthly temperature (red line) in the Liguria Region. Overall, the spatial variability of the temperature is about 1.6 °C. The month with the highest variability is December (standard deviation, σ, equal 2.0 °C), whereas June is the month with the lowest one (σ = 1.4 °C). The spatial variability of the precipitation is 30.3 mm. July and November are the months with the lowest (σ = 13.6 mm) and the highest (σ = 51.4 mm) spatial variability, respectively.

The SWS values were calculated according to the following equation (Nyvall, 2002):

$$SWS = RD \times AWSC \quad (4)$$

where RD is the crop rooting depth (m), varying from "shallow" (0.45 m for leaf vegetables) to "deep" (1.20 m for tree fruits), and AWSC is the available water storage capacity of the soil (mm/m), which ranges from 83 mm/m (sand textural class) to 208 mm/m (silt textural class). Since RD and AWSC are not available for the entire study area, we calculated all the possible SWS values (Table 1) with equation (4).

Using the MSWB model, we estimated the average monthly potential evapotranspiration $\overline{PET}_m$ (green curve in Fig. 4b) and

Table 1
Soil water storage (SWS, mm) values as a result of crop rooting depth (RD) and available water storage capacity (AWSC) values.

		RD (m)			
		0.45 Shallow	0.60 Medium Shallow	0.90 Medium Deep	1.20 Deep
AWSC (mm/m)	208 - Silt loam	94	125	187	250
	200 - Clay loam	90	120	180	240
	175 - Loam	79	105	158	210
	142 - Fine sandy loam	64	85	128	170
	125 - Sandy loam	56	75	113	150
	100 - Loamy sand	45	60	90	120
	83 - Sand	37	50	75	100

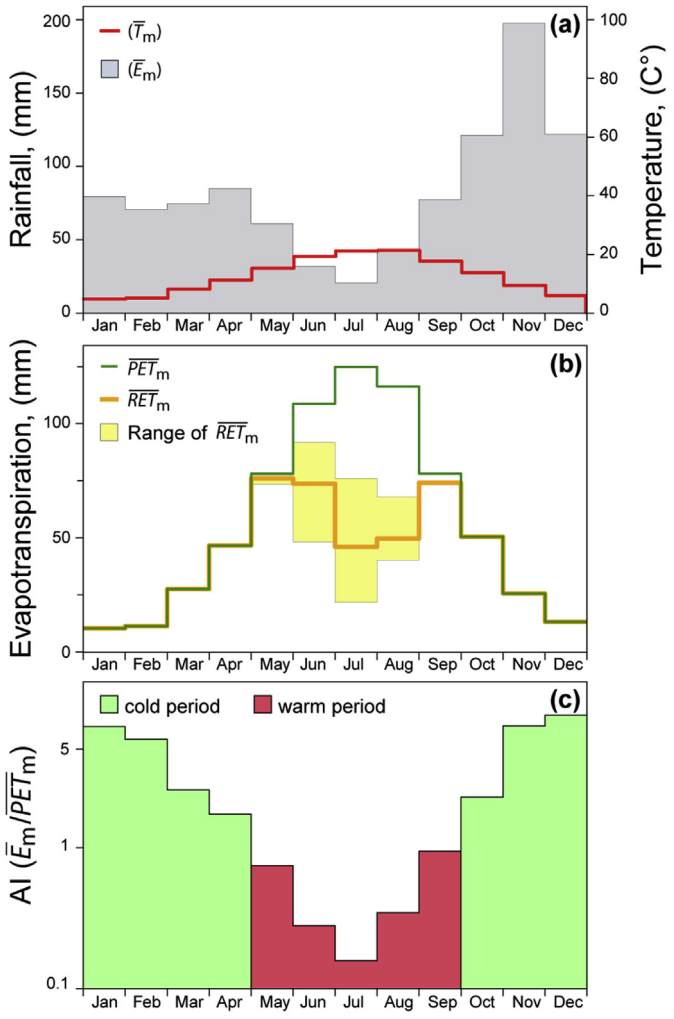

Fig. 4. (a) Average monthly rainfall (grey histogram) and the average monthly temperature (red line) in the study area. (b) Average monthly potential evapotranspiration, $\overline{PET}_m$, (green curve); mean value of $\overline{RET}_m$ (orange curve) as a function of the month and the range of variability of $\overline{RET}_m$ (yellow shaded area). (c) Warm (red portion of the histogram) and cold (green portion of the histogram) periods. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the average monthly real evapotranspiration $\overline{RET}_m$ corresponding to each SWS value. The yellow shaded area in Fig. 4b represents the range of variability of $\overline{RET}_m$ obtained varying SWS. The orange curve in the figure is the mean value of $\overline{RET}_m$. As expected, $\overline{RET}_m$ decreases in the warm period when the rainfall (grey histogram in Fig. 4a) decreases. We found the C_W and C_C periods using the average monthly Aridity Index, $\overline{AI}_m$, (Barrow, 1992):

$$\overline{AI}_m = \frac{\overline{E}_m}{\overline{PET}_m} \quad (5)$$

Specifically, C_W is the period when the soil exhibits a water deficit and is mostly dry ($\overline{E}_m$ less than $\overline{PET}_m$ and $\overline{AI}_m \leq 1$), and is from May to September (red portion of the histogram in Fig. 4c). Conversely, the C_C period is from October to April when $\overline{AI}_m > 1$ (green portion of the histogram in Fig. 4c).

Finally, we obtained $R = 2.4 \pm 0.3$ calculating the ratio of the total amount of $\overline{RET}_m$ between May and September (C_W period) and between October and April (C_C period). The uncertainty of *RET ratio* derives from the variability of $\overline{RET}_m$ (yellow shaded area in Fig. 4b). As a result, in Liguria R ranges broadly from a value between 2 and 3, which is close to the heuristic value $R = 2$ adopted by Melillo et al. (2015) in Sicily. In the following analysis, we reported the reconstructed rainfall events, rainfall conditions responsible for the failures and the relative rainfall thresholds assuming $R = 2$, which corresponds to a minimum dry interval of 48 h and 96 h in the C_W and in the C_C period, respectively. Then, we assumed $R = 3$ which corresponds to a minimum dry interval of 48 h in the C_W and 144 h and in the C_C period and we compared the distributions of the rainfall events and of the rainfall conditions, and the obtained thresholds.

5. Results

Adopting CTRL–T and using information presented in the previous sections, rainfall events and empirical rainfall thresholds at several exceedance probabilities were determined. The empirical cumulative distribution function (ECDF) of the duration D and of the cumulated rainfall E are shown in Fig. 5 for the *RE* (green curves), *MRC* (purple curves) and *MPRC* data set (orange curves) obtained by CTRL–T assuming $R = 2$. The corresponding statistics are reported in Table 2. For each data set, the table contain the total number, the minimum, and the maximum values and the first, second, and third quartiles (Q_1, Q_2, and Q_3) of the reconstructed rainfall events (*RE*) and rainfall conditions (*MRC* and *MPRC*).

Fig. 5a reveals that more than 70% of the reconstructed-rainfall events (*RE*, green curve) have a duration $D < 100$ h and 25% have $D \leq 6$ h (Table 2). The shortest values of D are observed in the *MPRC* data set (orange curve). Fig. 5b shows the ECDF for the *RE*, *MRC* and *MPRC* data sets. In particular, the median of the cumulated rainfall E for *RE* is lower (<20%) than that for *MRC* and *MPRC* (Table 2) whereas the duration D is comparable for the three data sets. As a consequence, the triggering conditions in the *MRC* and *MPRC* data sets exhibit generally a higher mean rainfall intensity I (E/D) (Fig. 5c), representing the most severe conditions reconstructed for the single landslide. The same statistics indicates that the duration D of the *MRC* are generally longer than that of *MPRC*, while the cumulated rainfall E is substantially the same (Table 2).

The delay between the rainfall ending time and the landslide occurrence time is null (0 h) for the majority of *MPRC* (286). The delay is ≤ 3 h for 69 *MPRC* and is between 4 and 48 h for the remaining 85 *MPRC*.

Using the *MRC* and *MPRC* data sets, and adopting the method proposed by Brunetti et al. (2010) and Peruccacci et al. (2012), we determined objective cumulated event rainfall – rainfall duration (*ED*) thresholds, and their associated uncertainties, for the study

Table 2
Statistics of the *RE*, *MRC* and *MPRC* data set obtained by CTRL–T.

Data set	#	Duration D (h)					Cumulated rainfall E (mm)				
		min	Q_1	Q_2	Q_3	max	min	Q_1	Q_2	Q_3	max
RE	43191	1	6	29	96	2005	1.1	5.4	18.2	57.6	1384.4
MRC	801	2	26	60	130	517	6.8	63.2	109.0	184.8	526.7
MPRC	440	2	19	34	62	274	13.2	63.6	113.7	181.5	526.7

area (Fig. 6).

Fig. 6a shows in log-log coordinates, the 801 (D_L,E_L) pairs relative to the *MRC* data set (purple dots), together with the *ED* frequentist threshold at 5% exceedance probability ($T_{5,MRC,R2}$, in purple; see equation in Table 3) and its associated uncertainty (shaded area). Analogously, Fig. 6b shows the 440 (D_L,E_L) pairs for the *MPRC* data set (orange dots), the 5% *ED* frequentist threshold ($T_{5,MPRC,R2}$, in orange; see equation in Table 3) and the relative uncertainty (shaded area). Fig. 6c shows the same thresholds in linear coordinates and in the range $1 \le D \le 120$ h.

We note that $T_{5,MRC,R2}$ is higher than $T_{5,MPRC,R2}$ for durations $D < 10$ h; conversely, for longer durations, $T_{5,MPRC,R2}$ is getting higher due to the greater angular coefficient. The mean value of the difference among the cumulated rainfall E for the two thresholds is 7 mm (12%) for $1 \le D \le 300$ h. In particular, the E differences varies from values below 2 mm (4%) for $D = 24$ h to values that do not exceed the 12 mm (16%) for $D = 300$ h.

To evaluate how R affects the reconstruction of the rainfall events, rainfall conditions, and the thresholds, we conducted a sensitivity analysis assuming $R = 2$ and $R = 3$ (solid and dashed

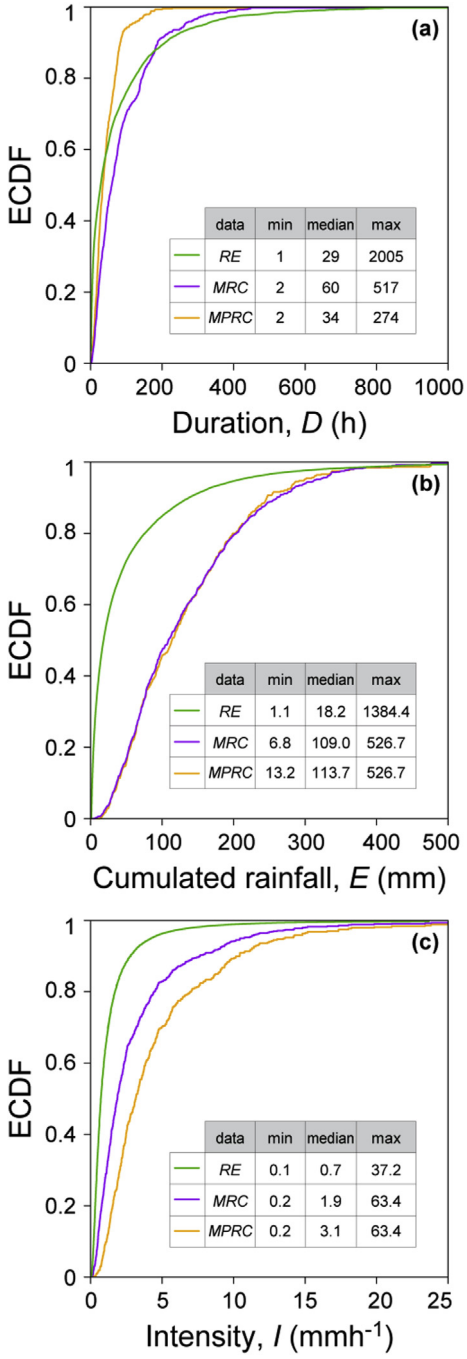

Fig. 5. ECDF curves of the duration D (a), cumulated rainfall E (b) and rainfall mean intensity I (c) of the reconstructed rainfall events (green curves), multiple rainfall conditions (purple curves) and maximum weight rainfall conditions (orange curves) obtained by CTRL–T assuming $R = 2$. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

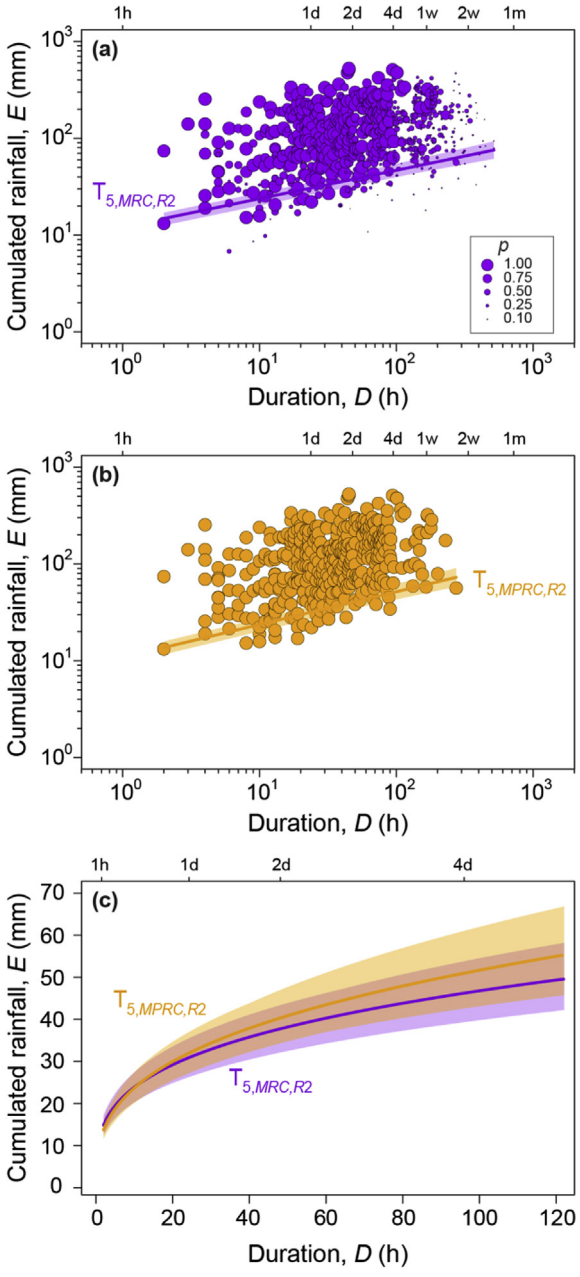

Fig. 6. Cumulated event rainfall E (mm) vs. rainfall duration D (h) conditions that have resulted in landslides in reconstructed by CTRL–T using (a) the *MRC* and (b) the *MPRC* data set. Purple and orange curves are the corresponding 5% thresholds, $T_{5,MRC,R2}$, $T_{5,MPRC,R2}$ (Table 1). In (a) frame, the size of the points (p probability) is proportional to the weight w of D_L,E_L pairs. The data are shown in log-log coordinates. (c) $T_{5,MRC,R2}$ and $T_{5,MPRC,R2}$ thresholds in the range $1\text{ h} \le D \le 120$ h, with associated uncertainty portrayed by shaded areas, in linear coordinates. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Equations of the rainfall thresholds for the possible initiation of landslides in Liguria for *MRC* and *MPRC* data sets varying the ratio *R* and the decay factor *k* and the weight *w*. *N* is the number of (D_L,E_L) pairs.

Data set	*N*	*R*	*w*	*k*	Threshold label	Threshold equation
MRC	801	2	$f(d, E_L, D_L)$	0.84	$T_{5,MRC,R2}$	$E = (12.1 \pm 1.8) \times D^{(0.29 \pm 0.04)}$
MRC	863	3	$f(d, E_L, D_L)$	0.84	$T_{5,MRC,R3}$	$E = (13.5 \pm 2.1) \times D^{(0.26 \pm 0.04)}$
MPRC	440	2	$f(d, E_L, D_L)$	0.84	$T_{5,MPRC,R2}$	$E = (10.8 \pm 1.7) \times D^{(0.34 \pm 0.04)}$
MPRC	440	3	$f(d, E_L, D_L)$	0.84	$T_{5,MPRC,R3}$	$E = (10.8 \pm 1.7) \times D^{(0.34 \pm 0.04)}$
MRC	916	2	$f(d, E_L, D_L)$	1.00	$T^{*}_{5,MRC,R2}$	$E = (8.6 \pm 0.9) \times D^{(0.43 \pm 0.04)}$
MRC[a]	916	2	$f(n)$	1.00	$T^{**}_{5,MRC,R2}$	$E = (7.3 \pm 1.1) \times D^{(0.44 \pm 0.04)}$

[a] Melillo et al. (2016).

curves, respectively, in the subsequent figures). For the purpose, we compared the ECDF of *D* and *E* for *RE*, *MRC*, *MPRC* (Fig. 7). In addition, we investigated the differences between the 5% thresholds for *MRC* and *MPRC* (Fig. 8).

Below Q_2 (i.e. $D < 30$ h), we found small differences (3%) between the ECDF curves of *D* for the *RE* data sets (green curves in Fig. 7a). The differences slowly increase up to 25% at Q_3 ($D \leq 100$ h). Analogously, in terms of *E*, the differences are between 5% and 10% from Q_2 to Q_3, in the range $20 \leq E \leq 65$ mm (green curves in Fig. 7b). For *MRC* the percentage difference in the ECDF of *D* is always 12% up to Q_3 (purple curves in Fig. 7a), while for *E* the ECDF differences do not exceed 5% (purple curves in Fig. 7b). Comparing the ECDF curves for the *MPRC*, we noted that they overlap, for both *D* and *E* (orange curves in Fig. 7a and b).

Log-log plot in Fig. 8 shows the 5% *ED* thresholds calculated using a *MRC* data set for $R = 2$ and $R = 3$. The obtained thresholds are substantially similar with differences in cumulated rainfall lower than 1 mm (1%) at $D = 24$ h. The maximum difference is 6 mm (7%) for the longest duration ($D = 517$ h). The 5% thresholds for *MPRC* are coincident (orange curves in Fig. 8).

6. Discussion

6.1. Algorithm improvements

CTRL–T aims at reducing the computational time of the single steps, from the rainfall event reconstruction to the threshold definition. The actions performed by CTRL–T prove robust, and the implementation software do not suffer from computer storage or processing limitations. More in details, the algorithm improvements reduce the calculation time significantly in the: i) selection of the available rain gauges, ii) identification of the representative rain gauge, and iii) reconstruction and analysis of the rainfall conditions responsible for the failure.

As for the first two points, the long manual procedure to associate a failure to the closest rain gauge is now overcome by the automatic selection of all the available stations within a given distance (R_b) from the landslide. R_b is parameterized and can vary as a function of the density of the rain gauges. CTRL–T identifies automatically the representative rain gauge through a weight, which is a function of both geographical (distance between landslide and rain gauge) and hydrological (cumulated rainfall and duration of the rainfall event) characteristics, associated to the

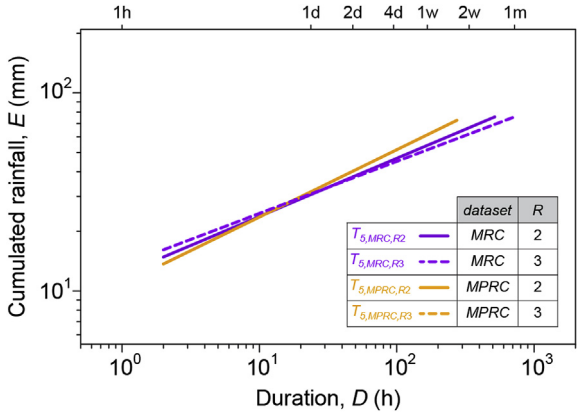

Fig. 8. *ED* thresholds for possible landslide occurrence, at 5% exceedance probability levels, reconstructed by CTRL–T using a *MRC* (purple curves) and *MPRC* (orange curves) dataset obtained assuming $R = 2$ (solid curves) and $R = 3$ (dashed curves), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

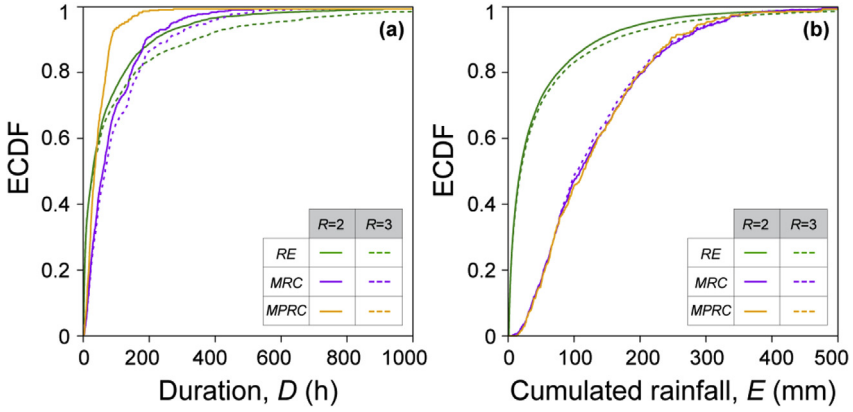

Fig. 7. Comparison between ECDF curves of the duration *D* (a) and cumulated rainfall *E* (b) assuming the two values of *R*. Solid and dashed curves represent the distribution for the three data sets assuming $R = 2$ and $R = 3$, respectively.

selected stations. With the proposed weight, the closest rain gauges are selected in almost 80% of the cases. Further improvements might be addressed to the weight evaluation by including orographic and geomorphological parameters, such as the elevation difference between landslide and rain gauge. The procedure implemented in CTRL–T allows running the algorithm also when the rainfall information is available as gridded rainfall maps (e.g., obtained through satellite data). In this case, each pixel of the map (e.g., 25 km × 25 km) is treated as a "virtual" rain gauge. Otherwise, the manual reconstruction of rainfall conditions using gridded rainfall maps would be time consuming.

Regarding the third point, the identification and selection of the rainfall conditions responsible for the landslide was previously left to the investigator and it was therefore subjective and time consuming. The introduction of the automatic and reproducible selection of rainfall conditions based on weights solves this drawback and tries to face one of the main sources of uncertainties in the threshold definition process. An additional improvement is obtained introducing the decay factor k. In the previous version, all the (D_L,E_L) pairs were considered equally probable for the activation of the landslide. However, conditions characterized by long durations and low mean intensity are unlikely to trigger a shallow landslide. The decay factor k damps the cumulated rainfall removing those rainfall condition with a small (<10%) rainfall increase compared to the previous condition.

In the new version of the algorithm, we add an action to calculate thresholds and their uncertainties adopting a frequentist method (Brunetti et al., 2010; Peruccacci et al., 2012). The procedure that begins from the reconstruction of the rainfall events and ends with the definition of thresholds is now faster than before and completely automatized, thus objective and reproducible. The structure of the algorithm also allows the implementation and/or integration of other threshold calculation methods.

Further details regarding CTRL–T, e.g., the required input files and the generated output files are available in Appendix A.

6.2. Parameter settings

As described in Melillo et al., (2015), in the previous version of the algorithm, the same values of the parameters were used to reconstruct the rainfall events from rainfall data recorded in all available rain gauges. This aspect represents a great limitation, given that the rain gauges located in wide areas might record differ rainfall regimes. In order to solve this problem, CTRL–T performs the reconstruction of the rainfall event using an input file (see Appendix A), containing the values of the selectable parameters, which can be adjusted for each rain gauge.

As stated in Section 4, we adopted the average monthly Aridity Index to establish the length of "cold" and "warm" periods. Then, we used the real evapotranspiration in the two periods to set the length of the two dry intervals (C_W and C_C). The calculation of these parameters is fundamental when using the algorithm in different seasonal and climatic contexts, and with various soil types. The procedure to determine C_W and C_C is still not automatized.

6.3. Rainfall events and thresholds

As reported in the previous section, the statistics of ECDF curves for *RE*, *MRC*, and *MPRC* data sets (Fig. 5) suggest that the triggering conditions selected in the *MRC* or *MPRC* data sets have a rainfall mean intensity I (E/D) higher than that of *RE*. Generally, to calculate the thresholds it is worth to use the most severe *MRC* (or *MPRC*) conditions instead of the whole rainfall event (*RE*) in order to avoid an underestimation of the rainfall responsible for the landslide. Note that only in few cases (1%), the (D,E) pairs for *MRC*, *MPRC* and *RE* coincide.

Fig. 6c and the equations of the two thresholds obtained for the *MRC* and *MPRC* data sets ($T_{5,MRC,R2}$ and $T_{5,MPRC,R2}$ in Table 3) reveal that the thresholds are statistically indistinguishable if considering the uncertainties associated with the curves (Fig. 6c). Since the number of rainfall conditions used to determine the thresholds is largely above 75, which is the minimum number to obtain stable α and γ parameters (Peruccacci et al., 2017), we acknowledge that the two curves are very similar despite the rather large uncertainties.

6.4. Sensitivity analysis of the R value

The two different values of the R parameter affect the ECDF curves of the reconstructed rainfall events (solid and dashed green curves in Fig. 7). While they are comparable in terms of cumulated rainfall E, they are dissimilar regarding the rainfall duration D. As the value of the minimum dry period in the cold season (C_C) increases from 96 h ($R = 2$) to 144 h ($R = 3$) the duration of the rainfall events increases and the total number of *RE* decreases. A higher value for the minimum dry period (needed to separate two subsequent rainfall events) makes more difficult the separation of events; consequently, the length of the rainfall event increases. The duration of the *MRC* is slightly longer when using $R = 3$ instead of $R = 2$, while the distribution of the cumulated rainfall remains unvaried (purple curves in Fig. 7). The ECDF curves for *MPRC* are overlapping (orange curves in Fig. 7) because the *MPRC* selected by CTRL–T in the two cases ($R = 2$ and $R = 3$) are coincident. This occurs if, for each landslide, the rainfall events reconstructed using $R = 2$ and $R = 3$ are the same or if the rainfall events for $R = 2$ are included in those reconstructed for $R = 3$ and the additional *MRC* related to $R = 3$ are not able to vary the selection of *MPRC*. However, differences are not so evident given that in the "cold" period C_C (November–March), 72% of the rainfall events reconstructed using $R = 2$ and $R = 3$ coincide. This means that these events are separated by at least a minimum dry period of 144 h.

Inspection of Fig. 8 reveals that the four thresholds obtained by using the *MRC* and *MPRC* data sets and the two R values do not differ much in the study area. In particular, thresholds for $R = 2$ ($T_{5,MRC,R2}$) and $R = 3$ ($T_{5,MRC,R3}$) are completely overlapping while the four thresholds, considering the uncertainties associated with them (Table 3), can be considered statistically indistinguishable. Therefore, the uncertainty introduced using different parameters (e.g. parameters obtained by the proposed *RET* analysis) is negligible and substantially does not affect the threshold calculation.

6.5. Threshold comparison

We compared the results obtained providing as input the same landslide and rainfall data sets to the old version of the algorithm (Melillo et al., 2016) and to the release proposed in this work. Fig. 9 shows the threshold $T^{**}_{5,MRC,R2}$ obtained using the previous version of the algorithm, and the thresholds $T^{*}_{5,MRC,R2}$ and $T_{5,MRC,R2}$ obtained using the new release of the algorithm and considering $k = 1$ and $k = 0.84$, respectively (see Table 3 for the threshold parameters). In particular, the $T^{*}_{5,MRC,R2}$ threshold has the same slope of the $T^{**}_{5,MRC,R2}$ and a higher intercept. The $T_{5,MRC,R2}$ threshold has a different slope and a range of duration shorter than the others, given that the use of the k decay factor decreases the durations of the rainfall conditions responsible for the failures. Moreover, using the k decay factor, the rainfall amount necessary to initiate landslides is lower for duration longer than 10 h (see $T_{5,MRC,R2}$ and $T^{*}_{5,MRC,R2}$ in Fig. 9).

The introduction of weights for the rainfall conditions and of a decay factor for the cumulated rainfall made the thresholds more similar to those defined by an expert investigator. In fact, the earlier

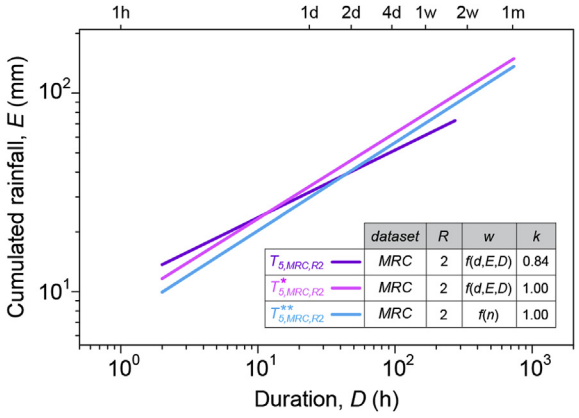

Fig. 9. *ED* thresholds for possible landslide occurrence, at 5% exceedance probability level, reconstructed by CTRL–T using a *MRC* dataset obtained by the old version of the algorithm (cyan curve) and new release (magenta and purple curves). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

version of the algorithm returned thresholds lower and steeper than those determined manually (Melillo et al., 2016). We maintain that CTRL–T reproduces better the reconstruction of the rainfall condition responsible for landslide triggering performed by an expert investigator.

7. Concluding remarks

The definition of reliable and accurate rainfall thresholds poses several critical issues and sources of subjectivity. The use of a standardized and automatized procedure for the reconstruction of the rainfall conditions responsible for failures and for threshold calculation is necessary for enhancing the objectivity and reproducibility of the thresholds, especially for thresholds to be used in landslide early warning systems.

In this work, we faced this problem and updated the algorithm proposed by Melillo et al. (2015, 2016), adding new features and ameliorations. CTRL–T exploits continuous rainfall measurements, and landslide information, to (1) reconstruct rainfall events; (2) select automatically the representative rain gauges; (3) identify multiple rainfall conditions responsible for the failure in terms of *D* and *E*, modelling the cumulated event rainfall; (4) attribute a probability to each rainfall condition; and (5) calculate probabilistic rainfall thresholds and their associated uncertainties.

The main innovations (novelties) concern (i) the automatic selection of the representative rain gauge that is chosen by analyzing the distance between the rain gauge and the landslide and the characteristics of the rainfall events recorded by the station, and (ii) the assignment of a probability to the single or multiple rainfall conditions responsible for the landslide. In addition, CTRL–T incorporates the procedures to calculate objective and reproducible thresholds developed by Brunetti et al. (2010) and Peruccacci et al. (2012).

We maintain that CTRL–T enhances the definition of empirical rainfall thresholds for landslide forecasting. In particular, the use of the algorithm standardizes and accelerates considerably the slow and tedious process of the reconstruction of the rainfall events and reduces the subjectivity inherent in the manual treatment of the rainfall and landslide data. This decreases the uncertainty associated with the definition of the rainfall conditions responsible for landslides and allows the possibility of a periodic and reproducible update of the thresholds, definitely. This has a positive impact on the application of thresholds in early warning systems for

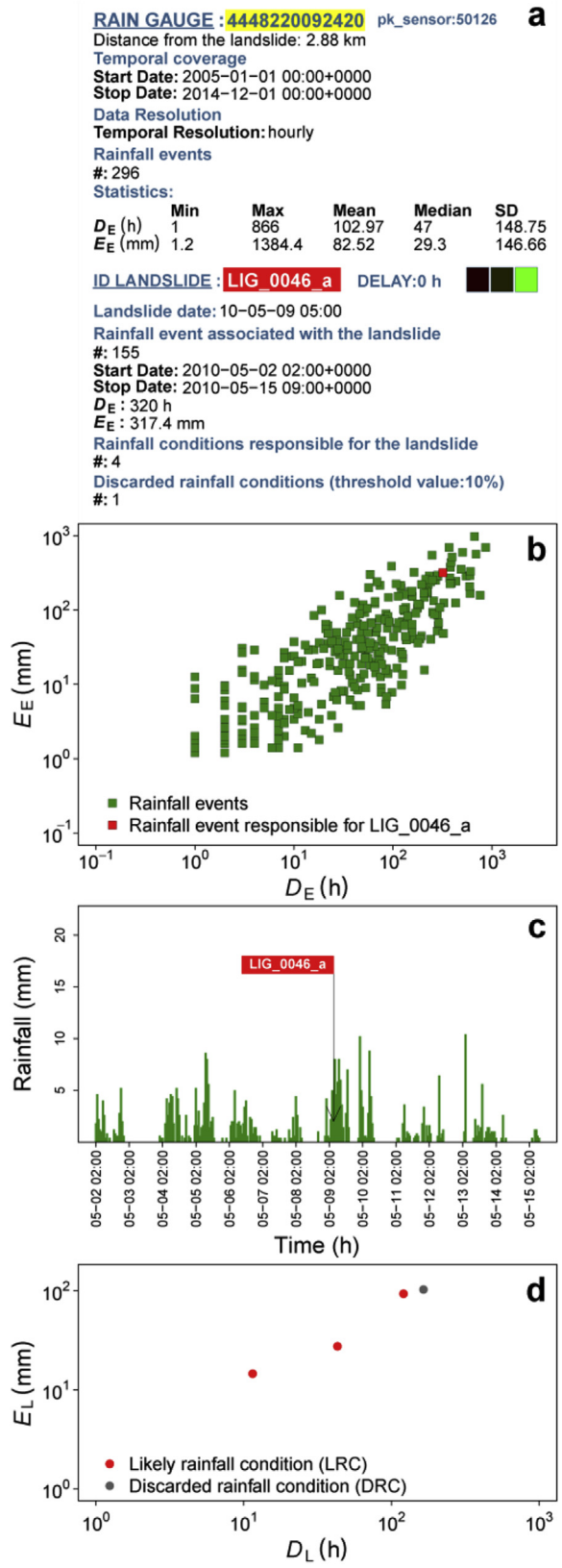

Fig. 10. Example of the output file named *Landslide_ID_project_ID_lan.pdf*. (a) Information about the descriptive statistics of the rain gauge; (b) distribution of the reconstructed (D_E,E_E) rainfall pairs in the logarithmic plane; (c) hourly rainfall measurements of the reconstructed rainfall event recorded by the representative rain gauge; (d) multiple (D_L,E_L) rainfall conditions responsible for the landslide in the logarithmic plane.

operational landslide forecasting.

Acknowledgements

This work was financially supported by the Italian national Department for Civil Protection (DPC) (Accordi di Collaborazione 2014, 2015, 2016), that also provided access to the rainfall database. We are grateful to Ivan Marchesini and Massimiliano Alvioli (CNR IRPI) who contributed to the calculation of the mean monthly temperatures used in this work. Massimo Melillo was supported by a grant from the Italian National Department for Civil Protection (Accordi di Collaborazione 2015, 2016). Stefano Luigi Gariano was supported by a grant from NERC in the framework of the LANDSLIP project (grant number NERC/DFID NE/P000681/1). Anna Roccati was supported by a grant from the Environment Department of the Liguria Region (Accordi di Collaborazione 2013, 2014). We thank the editor and four anonymous reviewers for their criticisms and comments that have helped us to improve the paper.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.envsoft.2018.03.024.

Appendix A. Software and data

We implemented CTRL–T (Calculation of Thresholds for Rainfall-induced Landslides–Tool) using the R open-source software for advanced statistical computing and graphics, release 3.3.3 (http://www.r-project.org). CTRL–T was finalized in June 2017 and should work on computers with at least 4 GB of RAM. The code can be downloaded free of charge at the following website: http://geomorphology.irpi.cnr.it/tools/rainfall-events-and-landslides-thresholds/ctrl-algorithm/ctrl-code/CTRL_code.R/. The downloadable executable file is 79 KB. In addition, an example of the input files required by the algorithm can be downloaded free of charge at the following website http://geomorphology.irpi.cnr.it/tools/rainfall-events-and-landslides-thresholds/ctrl-algorithm/input-demo/INPUT.zip/. The downloadable file is 37.55 MB.

The algorithm is divided into three main blocks (Fig. 1), which perform i) the reconstruction of rainfall events, ii) reconstruction of the rainfall conditions responsible for the landslide and iii) the definition of the rainfall thresholds, respectively.

Input files

Here, we describe the information provided by the input text files (boxes labelled with "*INPUT*" in the left upper part of the logical

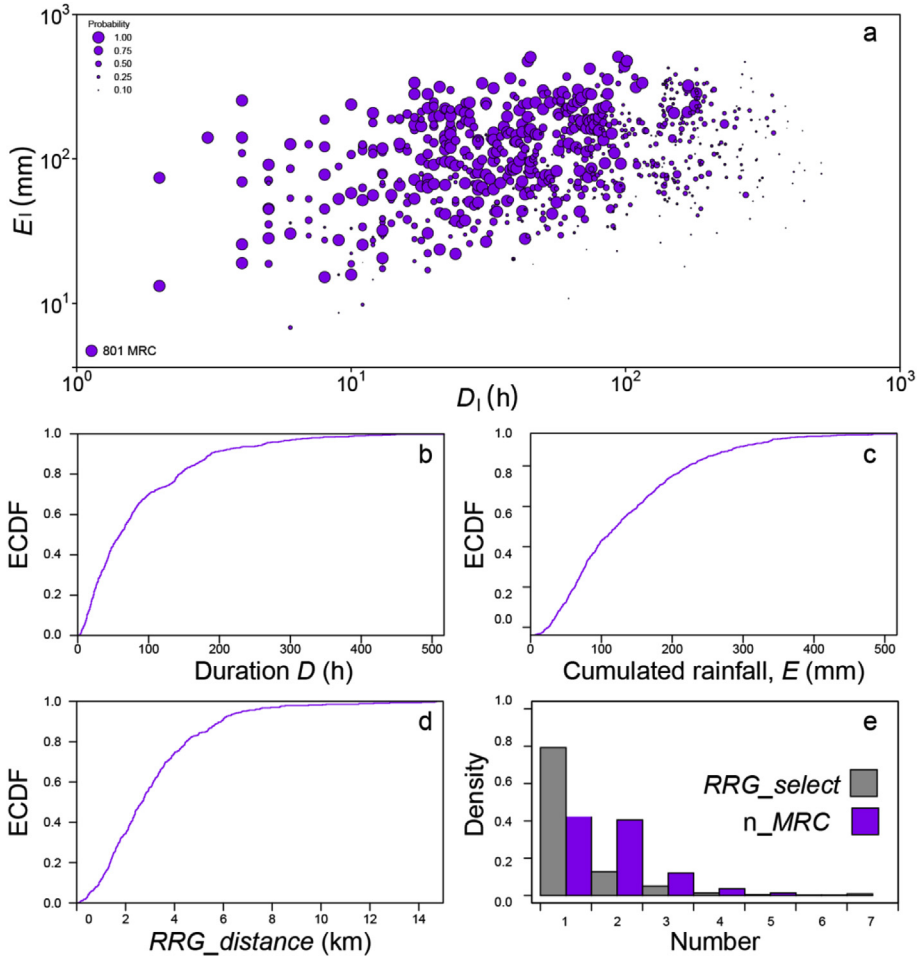

Fig. 11. Example of the output file named *Landslide_ID_project_ID_lan.pdf*, for the *MRC* data set (first page): (a) distribution of the (D_L,E_L) pairs, in log-log coordinates (purple dots); (b) ECDF of *D*; (c) ECDF of *E*; (d) ECDF of *RRG_distance*; (e) marginal distributions of the rank of representative rain gauges based on the distance from the landslides within the circular buffer, *RRG_select* (grey bars) and of the number of multiple rainfall conditions, *n_MRC* (purple bars). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

scheme in Fig. 1). One input file (*table_of_raingauges.csv*) contains information on the rain gauges (*cod_area*, *longitude*, *latitude*, *pk_sensor*, *ID_rain_gauge*) and on the parameters used by the algorithm to reconstruct the rainfall events (*gs*, *p1_c*, *p1_w*, *p2_c*, *p2_w*, *p3*, *p4_c*, *p4_w*, *sws*, *ews*) as described in Melillo et al., (2015). In particular, *cod_area* is an identification code of the administrative region where the sensor is located, *longitude* and *latitude* are the coordinates (expressed in decimal degrees), and *ID_rain_gauge* is a unique numeric code derived from the rain gauge geographic coordinates. A series of input files named *Time-Series_Sensor_ID_rain_gauge.csv* contain the time stamp and the hourly rainfall measurements of each sensor. An additional input file named *landslides. csv* contains information on the documented landslides. In particular, each record reports the: (i) *ID_project*, which identifies the rainfall event; (ii) *ID_lan*, which gives the temporal order of the failures associated with the single rainfall event (i.e., a is the 1st landslide, b the 2nd, etc.); (iii) *class_number*, which indicates the landslide multiplicity; (iv) *class_type*, which is the failure type; (v) *longitude* and *latitude* of the landslide; (vi) *geo_acc*, which is the level of mapping accuracy (Peruccacci et al., 2012); (vii) failure occurrence *date* (day, month, year and time); and (viii) *date_acc*, which is the accuracy of the occurrence date (T_1, T_2, and T_3). T_1 includes landslides for which the exact time of occurrence is known. T_2 and T_3 are used when the part of the day or the day of occurrence are known, respectively.

Output files

The three blocks of the algorithm (boxes in the right part of the logical scheme in Fig. 1) produce different output files.

The first block performs the reconstruction of the rainfall events and generates the folder *Reconstructed rainfall events*, which contains an output file named *Rainfall events. csv*, listing the reconstructed rainfall events. In particular, for each record *ID_rain_gauge*, is the rainfall station code, *index_pos1* and *index_pos2* are indexes related to the rainfall series of each rain gauge, *RE_start_date* and *RE_end_date*, are the starting and the ending date of each rainfall event. The remaining fields are the rainfall duration (*D_E*), the cumulated rainfall (*E_E*), the rainfall mean intensity (*I_E*), the maximum hourly rainfall (*IP_E*), the maximum cumulated rainfall in 24 h (*Emax24_E*), and a class (*A_class*), from 1 to 6, related to the rainfall event classification of Alpert et al. (2002). After the reconstruction of rainfall events, the algorithm calculates the maximum hourly rainfall and the maximum cumulated rainfall in 24 h. This allows to check, and possibly remove anomalous rainfall events.

The second block reconstructs the multiple conditions (*MRC*) likely responsible for the failures and generates the folder *Reconstructed rainfall conditions*, which contains three outputs. The first output is the folder *Individual files* containing a single file for each landslide named *Landslide_ID_project_ID_lan.pdf*. This file contains, for each rain gauge included in a buffer centered in the landslide location (see section 2.1), the: (i) information about the descriptive statistics of the rain gauge (Fig. 10a).; (ii) distribution of the reconstructed (D_E,E_E) rainfall pairs in the logarithmic plane (Fig. 10b); (iii) hourly rainfall measurements of the reconstructed rainfall event recorded by the representative rain gauge (Fig. 10c); (iv) multiple (D_L,E_L) rainfall conditions responsible for the landslide in the logarithmic plane (Fig. 10d). The second output is a text file named *Processing Summary report. txt*. This file contains the statistics about the total number of the analyzed, reconstructed, and discarded landslides. We produce a list of the discarded landslides, labelled with the *ID_project* attribute. The third output is a file named *MRC. csv*. In this file, *ID_project* identifies the rainfall event with landslide(s), *ID_lan* lists chronologically the landslides triggered by the same rainfall event, *date* is the failure occurrence time, *RRG_select* is the rank of the representative rain gauge based on the distance from the landslide within the circular buffer, and *RRG_distance* is the distance between the representative rain gauge and the landslide locations. The file lists also information on: the rainfall duration (*D_L*), the cumulated event rainfall (*E_L*), the

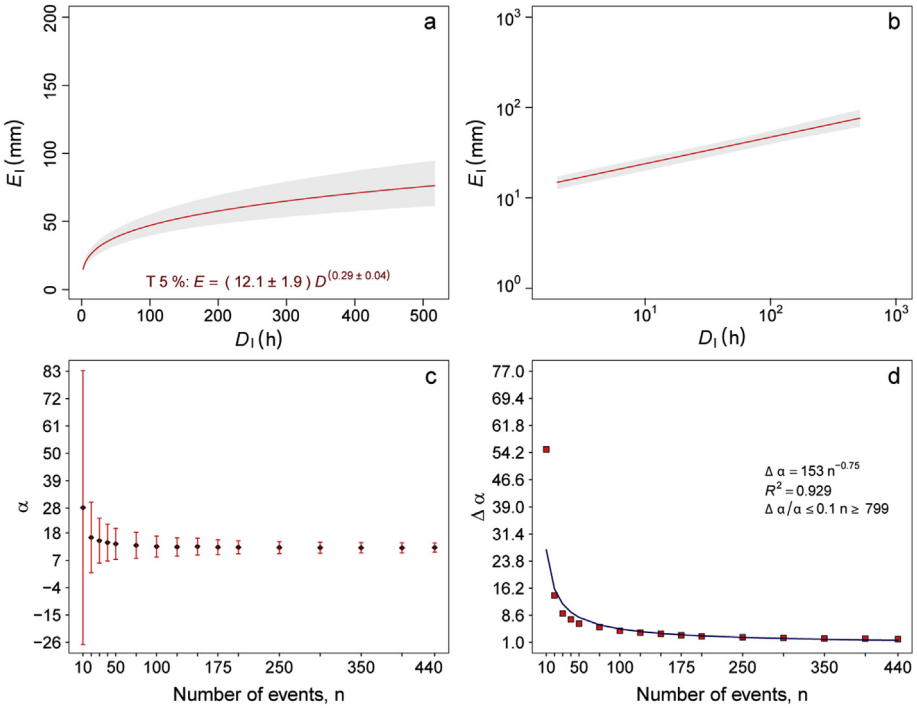

Fig. 12. Example of the output file named *Landslide_ID_project_ID_lan.pdf* (second to seventeenth pages): thresholds at 5% exceedance probability, in (a) linear and (b) log-log coordinates; variation of the values of (c) the α parameter and (d) the related uncertainty Δα.

rainfall mean intensity (*I_L*), the rain gauge code (*ID_rain_gauge*), the number of rainfall events (*REN*) associated with the *ID_rain_gauge*, the maximum hourly rainfall (*IP_L*), the maximum cumulated rainfall in 24 h (*Emax24_E*), a class (*A_class*, from 1 to 6) according to Alpert et al. (2002) rainfall event classification, a binary value (*ms_flag*) that indicates if the currently rainfall condition have a maximum value of the score *w* (*n_MRC* = 1), and the number of multiple rainfall conditions (*n_MRC*).

The third block defines the rainfall thresholds using *MRC* and *MPRC* data sets (see Section 2) and generates the folder *Rainfall thresholds*, which contains four outputs representing the results of the analysis. For each subset, the algorithm generates two distinct files. The first files (*boot_MRC*.csv or *boot_MPRC*.csv) contain information on the parameters (α and γ) that define 16 threshold curves, at different exceedance probabilities (from 0.005 to 50%). For each threshold, six fields are specified, i.e.: the name of the threshold parameters (*variable*), the corresponding exceedance probabilities (*probability*), the mean value (*mean*) obtained through the bootstrap process (Peruccacci et al., 2012), the standard deviation value (*sigma*) and the extreme values (*min*, *max*). The second files (*boot_MRC*.pdf or *boot_MPRC.pdf*) contain, in the first page: the distribution of the (D_L,E_L) pairs, in log-log coordinates (purple dots in Fig. 11a); the empirical cumulative distribution functions (ECDF) of *D* (Fig. 11b), *E* (Fig. 11c) and *RRG_distance* (Fig. 11d). Moreover, the marginal distribution of the representative rain gauge ranking, based on the distances between all rain gauges within the circular buffer and the landslide (*RRG_select*), and of the number of multiple rainfall conditions *n_MRC* are also reported (grey and purple bars in Fig. 11e, respectively). Here we report as example the file related to the *MRC* data set; the one related to the *MPRC* data set is analogous. In the subsequent 16 pages, are reported the graphs related to: i) the thresholds at exceedance probabilities from 0.005 to 50%, in linear and log-log coordinates (Fig. 12a, b, respectively), and the variation of the values of the α parameter and the related uncertainty $\Delta\alpha$ (Fig. 12c, d, respectively).

References

Aleotti, P., 2004. A warning system for rainfall-induced shallow failures. Eng. Geol. 73, 247–265. https://doi.org/10.1016/j.enggeo.2004.01.007.

Alley, W., 1984. On the treatment of evapotranspiration, soil moisture accounting, and aquifer recharge in monthly water balance models. Water Resour. Res. 20, 1137–1149. https://doi.org/10.1029/WR020i008p01137.

Alpert, P., Ben-Gai, T., Baharan, A., Benjamini, Y., Yekutieli, D., Colacino, M., Diodato, L., Ramis, C., Homar, V., Romero, R., Michaelides, S., Manes, A., 2002. The paradoxical increase of Mediterranean extreme daily rainfall in spite of decrease in total values. Geophys. Res. Lett. 29 (11) https://doi.org/10.1029/2001GL013554, 31–1–31-4.

Alvioli, M., Baum, R.L., 2016. Parallelization of the TRIGRS model for rainfall-induced landslides using the message passing interface. Environ. Model. Software 81, 122–135. https://doi.org/10.1016/j.envsoft.2016.04.002.

Alvioli, M., Guzzetti, F., Rossi, M., 2014. Scaling properties of rainfall induced landslides predicted by a physically based model. Geomorphology 213, 38–47. https://doi.org/10.1016/j.geomorph.2013.12.039.

An, H., Viet, T.T., Lee, G., Kim, Y., Kim, M., Noh, S., Noh, J., 2016. Development of time-variant landslide-prediction software considering three-dimensional subsurface unsaturated flow. Environ. Model. Software 85, 172–183. https://doi.org/10.1016/j.envsoft.2016.08.009.

Barrow, C.J., 1992. World Atlas of Desertification (United Nations Environment Programme), London. ISBN: 0340555122, p. 69.

Berti, M., Martina, M.L.V., Franceschini, S., Pignone, S., Simoni, A., Pizziolo, M., 2012. Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. J. Geophys. Res. 117 https://doi.org/10.1029/2012JF002367. F04006.

Brunetti, M.T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., Guzzetti, F., 2010. Rainfall thresholds for the possible occurrence of landslides in Italy. Nat. Hazards Earth Syst. Sci. 10, 447–458. https://doi.org/10.5194/nhess-10-447-2010.

Crozier, M.J., 1999. Prediction of rainfall-triggered landslides: a test of the antecedent water status model. Earth Surf. Process. Landforms 24, 825–833. https://doi.org/10.1002/(SICI)1096-9837(199908)24:9<825::AID-ESP14>3.0.CO;2-M.

Di Matteo, L., Dragoni, W., Giontella, C., Melillo, M., 2011. Bilancio Idrico del Lago Trasimeno. In: Martinelli, A. (Ed.), Tutela Ambientale del lago Trasimeno, pp. 69–84. Perugia, Arpa Umbria, ISBN/ISSN: 9788890592003 (in Italian).

Dragoni, W., Cambi, C., Di Matteo, L., Giontella, C., Melillo, M., Valigi, D., 2015. Possible response of two water systems in central Italy to climatic changes. In: Moramarco, T., Barbetta, S., Brocca, L. (Eds.), Advances in Watershed Hydrology. Water Resources Publications, LLC, pp. 397–424. ISBN-13: 978-1-887-20185-8.

Galanti, Y., Barsanti, M., Giannecchini, R., D'Amato Avanzi, G., Benvenuto, G., 2017. Statistical methods for the assessment of rainfall thresholds for triggering shallow landslides: a case study. In: Mikoš, M., Casagli, N., Yin, Y., Sassa, K. (Eds.), Advancing Culture of Living with Landslides. WLF 2017. Springer, Cham, pp. 429–436.

Gariano, S.L., Brunetti, M.T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O., Vennari, C., Guzzetti, F., 2015. Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, Southern Italy. Geomorphology 228, 653–665. https://doi.org/10.1016/j.geomorph.2014.10.019.

Giannecchini, R., Galanti, Y., D'Amato Avanzi, G., 2012. Critical rainfall thresholds for triggering shallow landslides in the Serchio River Valley (Tuscany, Italy). Nat. Hazards Earth Syst. Sci. 12, 829–842. https://doi.org/10.5194/nhess-12-829-2012.

Guo, D., Westra, S., Maier, H.R., 2016. An R package for modelling actual, potential and reference evapotranspiration. Environ. Model. Software 78, 216–224. https://doi.org/10.1016/j.envsoft.2015.12.019.

Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.P., 2007. Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorol. Atmos. Phys. 98 (3), 239–267. https://doi.org/10.1007/s00703-007-0262-7.

Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.P., 2008. The rainfall intensity–duration control of shallow landslides and debris flows: an update. Landslides 5 (1), 3–17. https://doi.org/10.1007/s10346-007-0112-1.

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 (A), 230–245. https://doi.org/10.1016/j.jhydrol.2016.01.018.

Keefer, D.K., Larsen, M.C., 2007. Assessing landslide hazards. Sci 316, 1136–1138.

Lepore, C., Destro, E., Nikolopoulos, E.I., Zoccatelli, D., Creutin, J.D., Guzzetti, F., Borga, M., 2017. Physically based modeling of rainfall-triggered landslides: a case study in the Luquillo forest, Puerto Rico. Hydrol. Earth Syst. Sci. 17, 3371–3387. https://doi.org/10.5194/hess-17-3371-2013.

Marra, F., Destro, E., Nikolopoulos, E.I., Zoccatelli, D., Creutin, J.D., Guzzetti, F., Borga, M., 2017. Impact of rainfall spatial aggregation on the identification of debris flow occurrence thresholds. Hydrol. Earth Syst. Sci. 21, 4525–4532. https://doi.org/10.5194/hess-21-4525-2017.

Martelloni, G., Segoni, S., Fanti, R., Catani, F., 2012. Rainfall thresholds for the forecasting of landslide occurrence at regional scale. Landslides 9, 485–495. https://doi.org/10.1007/s10346-011-0308-2.

Melillo, M., 2009. Modellizzazione del Lago Trasimeno. PhD Thesis. Accessed November 2017 (in Italian). http://www.scribd.com/doc/187518230/Modellizzazione-Lago-Trasimeno.

Melillo, M., Brunetti, M.T., Peruccacci, S., Gariano, S.L., Guzzetti, F., 2015. An algorithm for the objective reconstruction of rainfall events responsible for landslides. Landslides 12 (2), 311–320. https://doi.org/10.1007/s10346-014-0471-3.

Melillo, M., Brunetti, M.T., Peruccacci, S., Gariano, S.L., Guzzetti, F., 2016. Rainfall thresholds for the possible landslide occurrence in Sicily (Southern Italy) based on the automatic reconstruction of rainfall events. Landslides 13 (1), 165–172. https://doi.org/10.1007/s10346-015-0630-1.

Nadim, F., Jaedicke, C., Smebye, H., Kalsnes, B., 2013. Assessment of global landslide hazard hotspots. In: Sassa, K., Rouhban, B., Briceno, S., McSaveney, M., He, B. (Eds.), Landslides: Global Risk Preparedness. Springer-Verlag, pp. 59–71. https://doi.org/10.1007/978-3-642-22087-6_4.

Nadim, F., Kjekstad, O., Peduzzi, P., Herold, C., Jaedicke, C., 2006. Global landslide and avalanche hotspots. Landslides 3 (2), 159–173. https://doi.org/10.1007/s10346-006-0036-1.

Naoum, S., Tsanis, I.K., 2003. Hydroinformatics in evapotranspiration estimation. Environ. Model. Software 18, 261–271. https://doi.org/10.1016/S1364-8152(02)00076-2.

Nikolopoulos, E.I., Crema, S., Marchi, L., Marra, F., Guzzetti, F., Borga, M., 2014. Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence. Geomorphology 221, 286–297. https://doi.org/10.1016/j.geomorph.2014.06.015.

Nyvall, J., 2002. Soil Water Storage Capacity and Available Soil Moisture. Ministry of Agriculture Food and Fisheries of British Columbia, p. 4. Order No. 619000–619001.

Peres, D.J., Cancelliere, A., 2014. Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach. Hydrol. Earth Syst. Sci. 18, 4913–4931. https://doi.org/10.5194/hess-18-4913-2014.

Peres, D.J., Cancelliere, A., 2016. Estimating return period of landslide triggering by Monte Carlo simulation. J. Hydrol. 541 (A), 256–271. https://doi.org/10.1016/j.jhydrol.2016.03.036.

Peres, D.J., Cancelliere, A., Greco, R., Bogaard, T.A., 2017. Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds. Nat. Hazards Earth Syst. Sci. Discuss. https://doi.org/10.5194/nhess-2017-328.

Petley, D., 2012. Global patterns of loss of life from landslides. Geology 40 (10), 927–930. https://doi.org/10.1130/G33217.1.

Peruccacci, S., Brunetti, M.T., Luciani, S., Vennari, C., Guzzetti, F., 2012. Lithological and seasonal control of rainfall thresholds for the possible initiation of landslides in central Italy. Geomorphology 139–140, 79–90. https://doi.org/10.1016/j.geomorph.2011.10.005.

Peruccacci, S., Brunetti, M.T., Gariano, S.L., Melillo, M., Rossi, M., Guzzetti, F., 2017. Rainfall thresholds for possible landslide occurrence in Italy. Geomorphology

290, 39–57. https://doi.org/10.1016/j.geomorph.2017.03.031.
Piciullo, L., Gariano, S.L., Melillo, M., Brunetti, M.T., Peruccacci, S., Guzzetti, F., Calvello, M., 2017. Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. Landslides 14, 995–1008. https://doi.org/10.1007/s10346-016-0750-2.
Pumo, D., Lo Conti, F., Viola, F., Noto, L.V., 2017. An automatic tool for reconstructing monthly time-series of hydroclimatic variables at ungauged basins. Environ. Model. Software 95, 381–400. https://doi.org/10.1016/j.envsoft.2017.06.045.
Rosi, A., Peternel, T., Jemec-Auflič, M., Komac, M., Segoni, S., Casagli, N., 2016. Rainfall thresholds for rainfall-induced landslides in Slovenia. Landslides 13, 1571–1577. https://doi.org/10.1007/s10346-016-0733-3.
Segoni, S., Rossi, G., Rosi, A., Catani, F., 2014. Landslides triggered by rainfall: a semi-automated procedure to define consistent intensity–duration thresholds. Comput. Geosci. 63, 123–131. https://doi.org/10.1016/j.cageo.2013.10.009.
Staley, D.M., Kean, J.W., Cannon, S.H., Schmidt, K.M., Laber, J.L., 2013. Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California. Landslides 10, 547–562. https://doi.org/10.1007/s10346-012-0341-9.
Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. Geogr. Rev. 38 (1), 55–94. https://doi.org/10.2307/210739.
Thornthwaite, C.W., Mather, J.R., 1957. Instructions and tables for computing potential evapotranspiration and the water balance. Publ. Climatol. Lab. Climatol. Drexel Inst. Technol. 10 (3), 185–311.
Thorp, J.M., 1986. Mesoscale storm and dry period parameters from hourly precipitation data. Atmos. Environ. 20 (9), 1683–1689. https://doi.org/10.1016/0004-6981(86)90115-0.
Vennari, C., Gariano, S.L., Antronico, L., Brunetti, M.T., Iovine, G., Peruccacci, S., Terranova, O., Guzzetti, F., 2014. Rainfall thresholds for shallow landslide occurrence in Calabria, southern Italy. Nat. Hazards Earth Syst. Sci. 14, 317–330. https://doi.org/10.5194/nhess-14-317-2014.
Vessia, G., Parise, M., Brunetti, M.T., Peruccacci, S., Rossi, M., Vennari, C., Guzzetti, F., 2014. Automated reconstruction of rainfall events responsible for shallow landslides. Nat. Hazards Earth Syst. Sci. 14 (9), 2399–2408. https://doi.org/10.5194/nhess-14-2399-2014.
Yates, D.N., 1996. WatBal: an integrated water balance model for climate impact assessment of river basin runoff. Int. J. Water Resour. D. 12 (2), 121–140. https://doi.org/10.1080/07900629650041902.