Elsevier Editorial System(tm) for Computers & Geosciences

Manuscript Draft

Manuscript Number: CAGEO-D-15-00434R2

Title: Mimic expert judgement through automated procedure for selecting rainfall events responsible for shallow landslide: a statistical approach to validation

Article Type: Research paper

Keywords: rainfall thresholds, shallow landslide, automated procedure, statistical tests, LANDTRAIN code

Corresponding Author: Dr. Giovanna Vessia, PhD

Corresponding Author's Institution: Università "G. d'Annunzio" di Chieti-Pescara

First Author: Giovanna Vessia, PhD

Order of Authors: Giovanna Vessia, PhD; Luca Pisano, PhD student; Carmela Vennari, PhD student; Mauro Rossi, Researcher; Mario Parise, Researcher

Abstract: This paper proposes an automated method for the selection of rainfall data (Duration, D, and Cumulated, E), responsible for shallow landslide initiation. The method mimics an expert person identifying D and E from rainfall records through a manual procedure whose rules are applied according to her/his judgement. The comparison between the two methods is based on 300 D,E pairs drawn from temporal rainfall data series recorded in a 30 days time-lag before the landslide occurrence. Statistical tests, employed on D and E samples considered both paired and independent values to verify whether they belong to the same population, show that the automated procedure is able to replicate the expert pairs drawn by the expert judgment. Furthermore, a criterion based on cumulated distribution functions (CDFs) is proposed to select the most related D,E pairs to the expert one among the 6 drawn from the coded procedure for tracing the empirical rainfall threshold line.

Response to Reviewers

All the minor revisions suggested by the editor are undertaken in this third version of the manuscript.
In addition, lines 264-284 have been rephrased in order to improve the clarity of the text, according to the following advice of `Reviewer #1:`

```
Authors added additional test which proved my previous doubts. The language
(not English but language at all) of that part requires improvement:
..differences of paired samples are NULL through student and rank tests...
..Table 3 shows the results of these hypotheses.... ? etc.
```

Highlights

- A code to calculate duration and cumulated rainfall related to occurred landslides
- Six duration and cumulated values are drawn by setting W and ΔT input variables
- An automated method for non-expert users to undertake expert tasks
- Automated and expert procedures are compared through statistical tests

Mimic expert judgement through automated procedure for selecting rainfall events responsible for shallow landslide: a statistical approach to validation

Vessia Giovanna[1,2], Pisano Luca[2,3], Vennari Carmela[2,4], Rossi Mauro[5], Parise Mario[2]

[1]*Department of Engineering and Geology, University "G. d'Annunzio" of Chieti-Pescara, Chieti Scalo, Italy*

[2]*Institute of Research for Hydrogeological Protection IRPI, National Research Council CNR, Bari, Italy*

[3]*Department of Biosciences and Territory, University of Molise, Pesche (IS), Italy.*

[4]*Department of Earth, Environmental and Resources Science, University of Naples Federico II, Napoli, Italy.*

[5]*Institute of Research for Hydrogeological Protection IRPI, National Research Council CNR, Perugia, Italy*

Correspondence e-mail: g.vessia@unich.it

Abstract. This paper proposes an automated method for the selection of rainfall data (Duration, D, and Cumulated, E), responsible for shallow landslide initiation. The method mimics an expert person identifying D and E from rainfall records through a manual procedure whose rules are applied according to her/his judgement. The comparison between the two methods is based on 300 D-E pairs drawn from temporal rainfall data series recorded in a 30 days time-lag before the landslide occurrence. Statistical tests, employed on D and E samples considered both paired and independent values to verify whether they belong to the same population, show that the automated procedure is able to replicate the expert pairs drawn by the expert judgment. Furthermore, a criterion based on cumulated distribution functions (CDFs) is proposed to select the most related D-E pairs to the expert one among the 6 drawn from the coded procedure for tracing the empirical rainfall threshold line.

Keywords: rainfall thresholds, shallow landslide, automated procedure, statistical tests, LANDTRAIN code

1) Introduction

Shallow landslides triggered by rainfalls are a common source of damage to infrastructures, casualties and interruption of functionality of transportation systems worldwide. For this reason early warning systems have been devised to predict their possible occurrence (Baum and Godt 2010; Rossi et al. 2012; Papa et al. 2013) commonly by means of empirical rainfall thresholds. These have been introduced since several decades (Wieczorek and Guzzetti, 1999; Berti et al. 2012; Peruccacci et al. 2012; Nikolopoulos et al., 2014; Segoni et al., 2014; Zhuang et al., 2015) and represent the minimum rainfall cumulated E or intensity I versus duration values D responsible for landslide initiation. D,E or D,I pairs have been selected according to several expert criteria implemented by manual procedures proposed for different geographic and geologic settings (see the works by Caine (1980), Innes (1983) from worldwide databases, Ceriani et al. (1994) and Bolley and Oliaro (1999) from the Italian Alps, Wilson et al. (1992) from Hawaii, Sandersen (1996) from Norway, Dahal et al. (2008) from Nepal). Recently, for the Italian territory an expert method has been proposed by Brunetti et al. (2010) and Peruccacci et al. (2012) to select D,I and D-E pairs, respectively, aimed at the identification of rainfall threshold for the initiation of shallow landslides. Through the expert method the latest Italian empirical rainfall threshold has been drawn using 2408 landslide events (Brunetti et al. 2015). This method has been used within the early warning system SANF (an acronym for national early warning system for rainfall-induced landslides) devised by the CNR-IRPI research group (Rossi et al. 2012) for the Italian Civil Protection Office (DPC). Within this research project, financially supported by the DPC, some attempts to implement automated procedure that mimic the expert judgement were addressed. Automated procedures are needed from local administrations which employ non-expert users to implement policies against hazards. At this aim, the automated procedure by Vessia et al. (2014) was implemented as a code in R language (R core team 2013). It enables non-expert users to retrieve multiple D,E or D,I pairs from an input datasets of shallow landslide events. Comparing expert and automated methods that, starting from an observed shallow landslide event, calculate the event rainfall likely to be responsible for the failure, is not a straightforward task. In fact, there is often the possibility of multiple choices for the rainfall event. This is typically reduced at only one selection in the expert method, based on the user experience. The expert user accomplishes the calculation of E and D through flexible judgement according to different rainfall patterns in different seasons or climatic conditions. On the other hand, an "automated method" should not depend on the operator, but rather be able to guarantee the repeatability of the working steps, even though with a typically lower degree of flexibility and systematic biases. In the following, the comparison between the preceding two types of methods that independently calculate D-E pairs is undertaken. The comparison between the D-E pairs is addressed through statistical tests. In detail, to make the comparison feasible, two conditions were met: (1) the same sample, consisting of 300 landslide events that occurred in Central-Southern Italy in the time span 2002-2012 was used; (2) the same criteria to define the time of the landslide onset and to derive its geographical location from the sources of information have been adopted. Main objective of the comparison is investigating the sample marginal distribution and moments of E and D to check whether both belong to the same population. If this is the case, the "automated method" can be considered to adequately reproduce the "expert" choice of the rainfall event that likely induced a shallow landslide. This means that the systematic bias introduced by a repetitive procedure does not heavily affect its calculations. In this regard, the automated method shows to be predictive like the expert method. To this end, statistical tests of hypotheses are used for paired and independent samples.

2) Methodological approach

The exhaustive description of the effects of rainfall events as inducing shallow landslides is not feasible due to many uncertain factors that are: 1) the landslide initiation time and its location, drawn by reliable sources of information, 2) the number of contemporary landslides and the time delays of multiple landslide initiation, 3) the contribution of evapotranspiration on the moisture conditions predisposing to landslide occurrence. As the sources of information are concerned, the most certain source is the direct observation of witnesses, better if they are landslide experts, which, however, is rarely the case. Thus, the main sources of information for scientists are newspapers or reports from fire fighters. These latter typically refer only to those landslides affecting the main transportation lines or urban centers. When single or multiple landslides occur outside urbanized areas they presumably go undetected. Furthermore, when a shallow landslide occurs along a transportation line, a spatial precision lower than 1km is easier to be acquired, although it strongly depends on the quality of the information sources. In these cases, it is cumbersome to associate a rain-gauge to this landslide. Shallow landslides with geographical precisions lower than $100km^2$ will not be taken into account in this article.

Concerning the evapotranspiration role within soil matrix, it reduces the wet condition of the surficial soil deposits. The magnitude of this contribution over the seasons in Mediterranean climate has been investigated during the last years by means of experimental studies, thus some assumptions can be posed.

Longobardi and Khaertdinova (2015) investigated the evapotranspiration fluxes during inter-storm periods at an experimental site in Southern Italy. Their measures pointed out that evapotranspiration affects the 10 cm depth much more than at 30 cm. Moreover, 15 days seem to be the time span needed for increasing 3-4% the water depletion at 30 cm, both during the wet and the dry seasons. Thus, at a seasonal scale, the rate of depletion appeared to be rather uniform throughout the year mostly for the deeper layers.

A laboratory testing in a climatic chamber has been performed by Cui et al. (2014) for investigating the desiccation process of clay with plasticity index PI=35%. Based on the measured volumetric water content and suction, the Longobardi and Khaertinova (2015) finding is confirmed; the drying process by evapotranspiration is not active at depth > 25 cm. The actual evaporation rate curve shows that the soil in the near surface zone (5 cm) become unsaturated after 5 days. Based on these investigations, it can be drawn that evapotranspiration cannot be considered influencing the landslide predisposing wet conditions. Thus it has been disregarded in the automated method. A detailed description of the "expert" and "automated" method is provided in Vessia et al. (2014), to which the reader is referred for further details. Here, the logical schemes of the two methods are summarized and shown in Fig. 1a,b, and the essential features of the methods are outlined as follows. An Online Resource is available at https://github.com/gvessia/LANDTRAIN.

The two procedures show some common characters, namely: 1) assuming the end time of the rainfall T_E as the time of the landslide onset; 2) searching for the start time of the rainfall event T_S; 3) using representative rain gauge records selected within a radius of 10 km from the landslide point; 4) dividing the spatial precision of the landslide events into 4 spatial precision classes: <1 km^2 (P1), 1-10 km^2 (P2), 10-100 km^2 (P3), 100-1000 km^2 (P4); 5) dividing the temporal precision into three classes: hourly (±1h), estimated portion of the day (±6h) and daily based (±24h) (Gariano et al. 2012). All these assumptions brought to the selection of the 300 landslide events analyzed hereafter.

The main differences between the two methods can be summarized as follows: the expert method (Fig. 1a) detects the starting time of the event rainfall by means of the visual insight of the cumulated rainfall over at most 20 days backwards T_E. The T_S is assumed as the first rainy hour after a dry period of ΔT=48-72h (during spring and summer) or ΔT=72-96h (during fall and winter).

On the contrary, the automated method (Fig. 1b) analyses a rainfall record of at least 30 days, going backwards from T_E, and looks for two time periods (ΔT=48h and 72h) along which the cumulates of three rainfall intensity series (with time windows W= 3, 6 and 12h) stop to increase. According to the findings in section 2.1, no differences in seasons are considered for the inter-storm time span, since evapotranspiration is disregarded. The T_S is assumed to be the first rainy hour after two periods ΔT=48 and 72h, characterized by very low rainfall intensity (ε<0.2mm/h). These two inter-storm periods are considered like the minimum time span to discriminate the end of a rainfall event and the beginning of a new one, independent of the previous. The two values try to catch four different rainfall patterns: short wet periods with short dry periods, short wet periods with long dry periods, long wet periods with short dry periods and long wet periods with long dry periods.

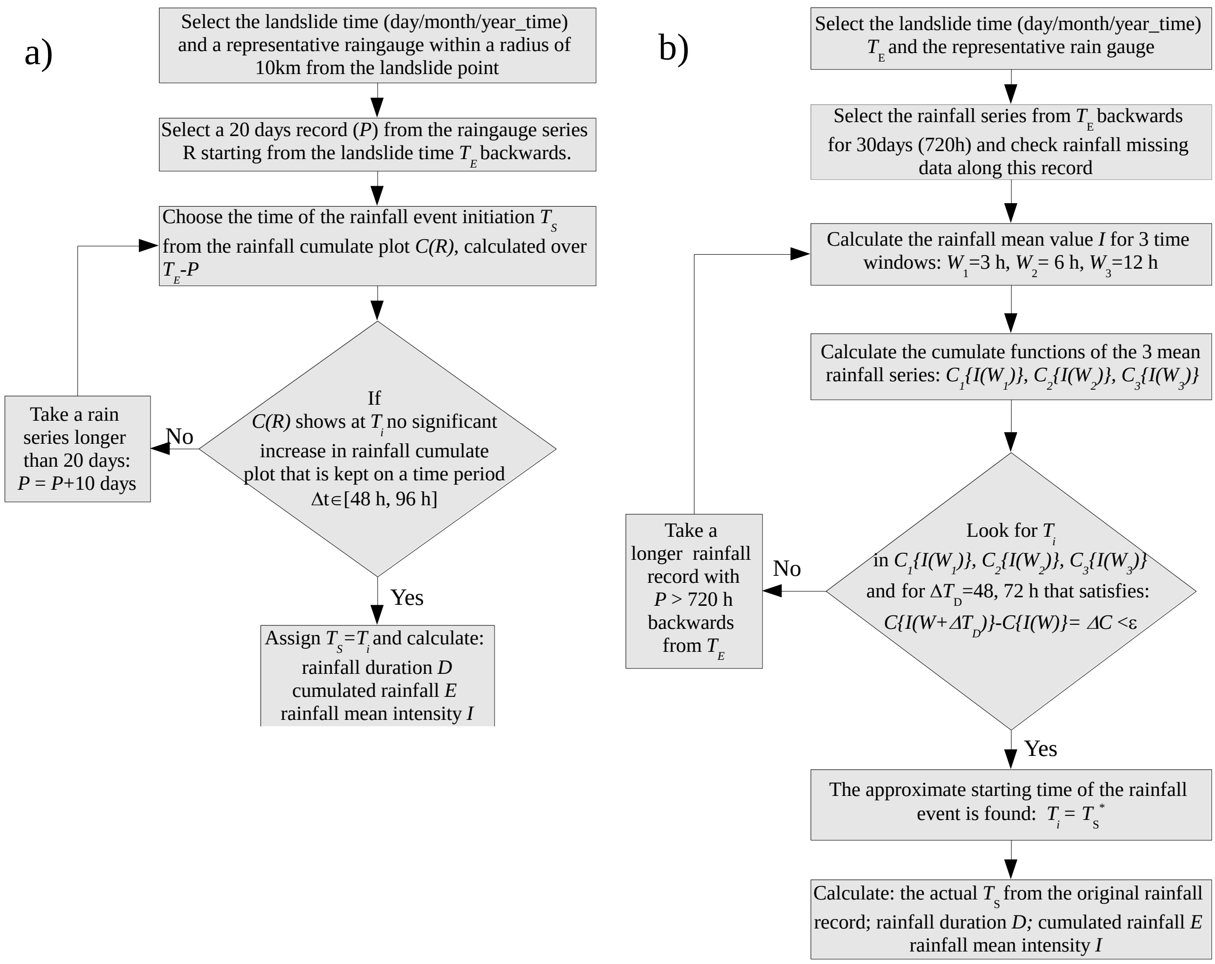

Fig. 1 Logical sequence of the working procedure to calculate (D,E) pairs from rain gauge rainfall records: a) the expert method; b) the automated method

The two methods provide different sets of (D,E) pairs for the same number of landslides: the expert method calculates one (D,E) pair for each landslide, whereas the automated method calculates 6 (D,E) pairs per landslide. The latter is due to the combinations of the following two parameters: 3 time windows W and 2 time periods ΔT. These six combinations simulate the expert judgment looking at the rainfall records in terms of temporal patterns. These latter change according to seasons, climatic zone and altitude. Simplifying, two models can be considered: i) the "convective rainfall", known as "storm" and characterized by high rainfall intensities occurred in limited portions of territory; and ii) the "frontal rainfall", which shows high cumulated rainfall values over long durations. These two models do not take into account hurricanes, or those "extreme" events that are commonly responsible for many effects at the ground surface.

3) Key features of the 300 shallow landslide samples

For the purpose of comparing the "automated" and the "expert" methods, 300 single shallow landslides located in Central-Southern Italy (Fig. 2a) have been selected from a larger database collected on the Italian territory by CNR-IRPI (Brunetti et al. 2015). The main properties of these 300 landslide events are shown in Fig. 2b-e. They have been extracted in GIS environment by using as base map a 20x20m DEM of Italy (Tarquini et al. 2007, 2012). The geographical precision in the location of the 300 landslides covers prevalently the classes P1 and P2, with 62% and 43%, respectively. Only a few landslides (4%) are characterized by P3 precision.

Figure 2b shows the temporal precision of the landslides: the majority (54%) are classified with hourly precision, while a presumed hour has been assigned to 31% of the landslides. Finally, 15% of the sample shows a daily precision.

The type of slope movements in the sample is illustrated in Fig. 2d. As it can be noted, five types of shallow landslides are considered. Nonetheless, for a significant part, corresponding to 34% of the sample, the mechanism is not known from the information source. Additionally, Fig. 2c-e show the climate and physiographic conditions related to the selected landslides. As concerns the climate, the Koppen-Geiger classification was used (Peel et al. 2007). The selected shallow landslides fall within four temperate climates, characteristic of the Mediterranean areas (Fig. 2c): Cs=Temperate sub-tropical; Cf6=Temperate Subcontinental; Cs7=Temperate Sub-coastal; Cs8=Temperate hot. Finally, modifying the Italian physiographic map "Carta Natura" (ISPRA 2013), at 1:250.000 scale, five physiographic units have been considered. The selected shallow landslides mostly occurred in (Fig. 2e) hill (61%), mountain (23%) and plain and coast (14%).

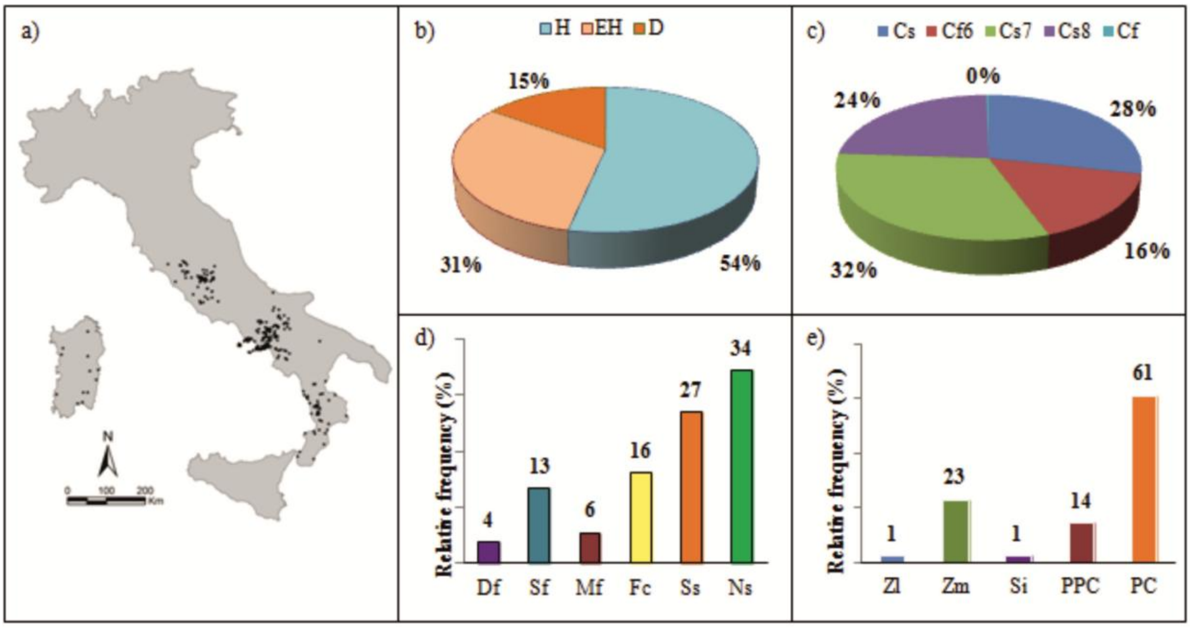

Fig. 2 A) Shallow landslide locations in Italy - B) Temporal accuracy (H=hourly; EH- estimated hour; D=daily) - C) Climate classification (Cs= Subtropical; Cf6= Subcontinental; Cs7= Sublittoral; Cs8= Warm Temperate) - D) landslide mechanism (Df=debris flow; Sf=soil flow; Mf=mud flow; Fc=fall; Ss=soil slip; Ns=non specified) - E) Physiographic units (Zl=Lacustrine zone/Lagoon; Zm=mountain; Si=Inter-mountain sectors; PPC= plain and coast; PC=hill)

This brief description of the sample of 300 shallow landslides enables to state that it is representative of the typical shallow landslides characterizing the central and southern portions of Italy. On these bases some general conclusions can be inferred from the following statistical analyses.

4) Statistical comparisons

4.1 The first glance for D-E pairs

The six different 300 (D,E) pairs drawn by the automated method and that calculated through the expert method are plotted in Figs. 3 a-f. The resulting mean threshold lines are also shown. As it can be noted, they show similar trends (derived by the ordinary least squares OSL method), although the coefficient of determination R^2 of the expert method is always higher than those related to the automated one.

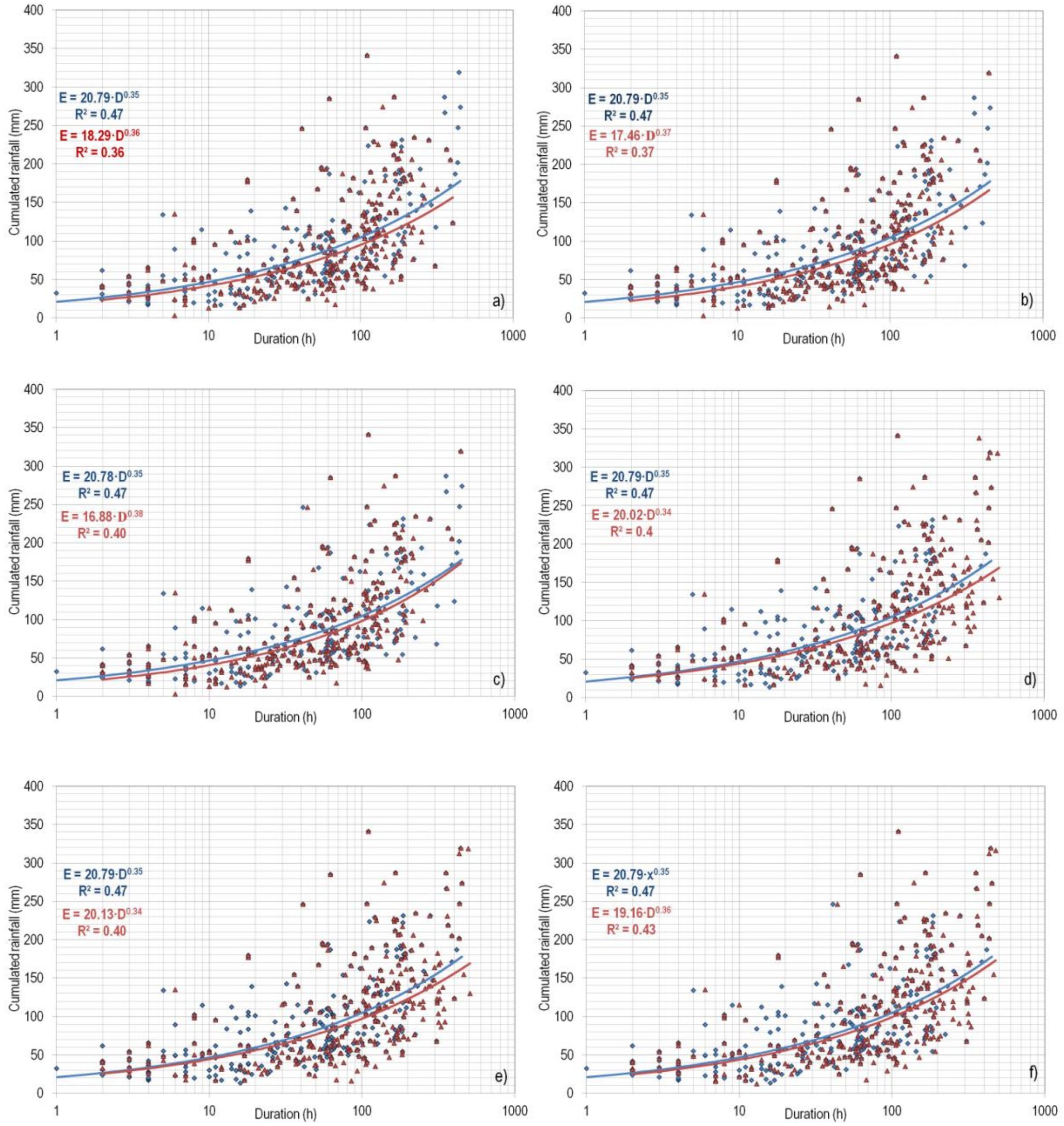

Fig. 3 Cumulated vs Duration rainfall mean threshold lines traced by means of the ordinary least squares method in abscissa log scale: blue pairs from expert method, red pairs from automated method. Combinations from automated method: a) 48x3; b) 48x6; c) 48x12; d) 72x3; e) 72x6; f) 72x12

It is worth noticing that the automated datasets show slightly different characters according to the parameter combinations *W* and *ΔT*. When *ΔT*=48h (Fig. 3a-c) the automated duration values fall between the expert extreme duration values. On the contrary, when *ΔT*=72h the automated dataset shows higher duration values than the expert ones, which are concentrated in the short and middle duration values. Moreover, the cumulated values from the automated method with *ΔT*=48h show lower values than the expert at middle duration values. These graphical evidences correspond to analytical differences in mean, median, standard deviation, minimum and maximum values of cumulated and duration samples (Table 1). As it can be argued from the cumulated sample, the expert estimator values fall within those related to the six automated samples. In particular, the 48 group shows lower cumulated estimators than the mean, median and minimum values of the expert method. The duration estimators, on the other hand, show higher mean and lower standard deviation values than the expert duration dataset. The 72 group is always higher than the expert estimator values, for both the cumulated and duration values. Thus D-E pairs from the 48 group resemble the expert ones rather than the 72 group. Carrying on the deterministic comparison, Table 2 shows the overlapping percentage of cumulated and duration values calculated by the two methods.

Table 1 Some relevant estimators of Cumulated and Duration samples drawn from the expert and the automated methods

CUMULATED	EXP	AUT 48x3	AUT 48x6	AUT 48x12	AUT 72x3	AUT 72x6	AUT 72x12
Mean	96.5	94.8	92.7	91.2	108.3	107.0	105.2
Median	78	74.6	70.1	69.1	93.3	93.3	91.1
Standard Deviation	63.4	60.0	61.0	61.1	67.8	66.4	66.7

Minimum	13.4	2.93	2.93	2.93	15.8	15.8	12.6
Maximum	341.2	341.2	341.2	341.2	341.2	341.2	341.2
DURATION							
Mean	83.8	88.3	81.3	76.1	128.5	123.9	113.9
Median	58	70	63	60	103.5	100.5	84.0
Standard Deviation	90.5	72.1	69.7	67.9	110.0	107.6	103.3
Minimum	1	2	2	2	2	2	2
Maximum	451	403	444	444	509	509	478

These overlapping percentages are calculated by considering two ranges of tolerance: ±10mm and ±10h for cumulated and duration values, respectively. They show very similar high values with slight differences between the 48 and 72 groups. In detail, the 48 group shows a 77-79% overlapping percentage for cumulated values, and a 64-69% for durations; the 72 group is characterized by lower overlapping percentages for both the cumulated and duration values. The reason for higher overlapping percentage in cumulated values than duration is due to the selection criteria: the automated method selects the rainfall event on the basis of the cumulated value; accordingly the duration is selected In doing this corresponding cumulated values can be related to different duration values according to selecting criteria used by the two methods.. Nonetheless, considering all the six samples from the automated method, the percentage of D,E overlapping rises to 82 and 90%, respectively. Precisely, the duration overlapping percentage is less than the cumulated one because, although the cumulated value is similar between the expert and one (or more than one) of the automated six combinations, the corresponding duration values can be more different than the cumulated values. As a matter of fact, ±10mm in cumulated value can differ much more than ±10h in duration value, depending on the pattern of the rainfall and the choices drawn from the expert judgement. For all these reasons, the overlapping percentages shown are considered satisfactory from a deterministic standpoint.

Table 2 Overlapping percentages of cumulated and duration values between the two methods: EXP=expert, AUT=automated

	Cumulated values (±10mm)	Duration values (±10h)
EXP –AUT 48x3	79%	64%
EXP –AUT 48x6	78%	66%
EXP –AUT 48x12	77%	69%
EXP –AUT 72x3	71%	58%
EXP –AUT 72x6	73%	61%
EXP –AUT 72x12	77%	66%
EXP – TOT AUT	90%	82%

4.2 Statistical tests for D-E pairs analyzed

In the study developed hereafter, some hypotheses are tested through parametric and non-parametric statistical tests on E rainfall cumulated values (and accordingly D values) drawn by not linear operations from different parsec of rain gauge recordings. It must be recalled that the two procedures use different criteria and rules to calculate E and accordingly D.

Moreover, the 300 values selected through expert judgement are drawn from many different experts. This implies the subjectivity of the application of common rules by different experts/scientists. Experts look at the rainfall hourly measures and sum them up backwards to find out the initiation of the rainfall event; the automated method looks at the rate of intensity variations in rainfall measures within different time windows. These two different procedures produce different E samples that cannot be considered paired yet, although referred to the same rain station related to the same landslide. This poses the question we are trying here to answer: “do the two methods generate statistically equivalent D,E samples to be used for rainfall threshold construction?”

In order to make a statistical comparison between the D,E datasets calculated by the two methods, statistical tests on the sample marginal distributions of D and E, and their mean values and variances, have been applied by using Statgraphics (StatPoint Technologies 2013). First, the tests for paired samples have been undertaken on the differences of paired E and D values calculated at the same rain stations from the two methods. Three null hypotheses have been checked: (1) the mean and (2) the median of differences of the paired samples are null; (3) the standard deviation of the paired samples is equal to 1. Student and rank tests were used for the verification of the first two hypotheses, while the χ^2 test was used for the third one. The null hypotheses have to be rejected if their probability P value is lower than 0.05. Table 3 shows the results of these tests: the three null hypotheses are rejected for almost all the paired samples. Hence D,E samples from the two methods cannot be considered belonging to the same population if they are taken as paired measures.

Table 3 Statistical comparisons of E and D paired samples: expert samples versus the 6 combinations drawn by the automated method. The null hypotheses are all checked at 95% of confidence. P values are listed in column.

Paired samples	**No differences between Mean Values (Student test)**	**No differences between medians (rank test)**	**No differences between Standard deviations (χ^2 test)**
Cumulated (expert vs 48x3)	0.35	0.0	0.0
Cumulated (expert vs 48x6)	0.06	0.03	0.0
Cumulated (expert vs 48x12)	0.02	0.12	0.0
Cumulated (expert vs 72x3)	0.0	0.0	0.0
Cumulated (expert vs 72x6)	0.0	0.0	0.0
Cumulated (expert vs 72x12)	0.0	0.0	0.0

Duration (expert vs 48x3)	0.32	0.0	0.0
Duration (expert vs 48x6)	0.58	0.0	0.0
Duration (expert vs 48x12)	0.09	0.07	0.0
Duration (expert vs 72x3)	0.0	0.0	0.0
Duration (expert vs 72x6)	0.0	0.0	0.0
Duration (expert vs 72x12)	0.0	0.0	0.0

Further, the three null hypotheses above have been also checked on the six paired samples from the automated method (calculated using three values of *W* and two values of ΔT). The following tests for multiple samples have been used: least significant difference (LSD) test for mean, Kruskal-Wallis test for median, and Leven test for variance values. The tests show that the three samples related to ΔT=48 belong to the same population. This is also true for the three samples related to ΔT=72. Conversely, the preceding null hypotheses fail to be tested when applied to the preceding 6 samples (grouping ΔT=48 and 72 samples). This implies that statistically meaningful differences in mean, median and variance values can be detected if samples are related to different ΔT values.

As a second step of this statistical study the 300 D-E pairs from the expert method are now taken as independent of the previous 6x300 pairs, since they were demonstrated not to belong to the same population as paired samples. This second condition is weaker than the preceding one but for the purpose of collecting D-E pairs of data (that is tracing empirical rainfall thresholds) it is acceptable. In fact, D-E pairs from the two methods are needed to generate a mean trend and variance that are “statistically” similar on the whole.

Table 4 Best fitting model distributions for D and E samples (null hypothesis checked through the Kolmogorov-Smirnov test at 95% confidence)

Sample	**Kolmogorov-Smirnov's test**	**P-value**

Cumulated expert	LogNormal	0.74
Cumulated 48x3	LogNormal	0.56
Cumulated 48x6	LogNormal	0.68
Cumulated 48x12	LogNormal	0.67
Cumulated 72x3	LogNormal	0.16
Cumulated 72x6	LogNormal	0.13
Cumulated 72x12	LogNormal	0.24
Duration expert	Weibull	0.46
Duration 48x3	Weibull	0.14
Duration 48x6	Weibull	0.31
Duration 48x12	Weibull	0.23
Duration 72x3	Weibull	0.59
Duration 72x6	Weibull	0.78
Duration 72x12	Weibull	0.78

Also in this second case, the statistical analyses are aimed at checking whether the D and E datasets drawn from the two methods belong to the same distribution and population. Frequency histograms of the E and D samples are separately studied and illustrated. Table 4 lists the best fitting model distributions of the samples estimated by means of the Kolmogorov-Smirnov goodness test. P-value, the probability value for the null hypothesis, checks the correspondence between the sample and the model probability distributions. This test verifies the null hypothesis at the 95% confidence level, meaning that P-value shall be higher than 0.05 when the null hypothesis is accepted.

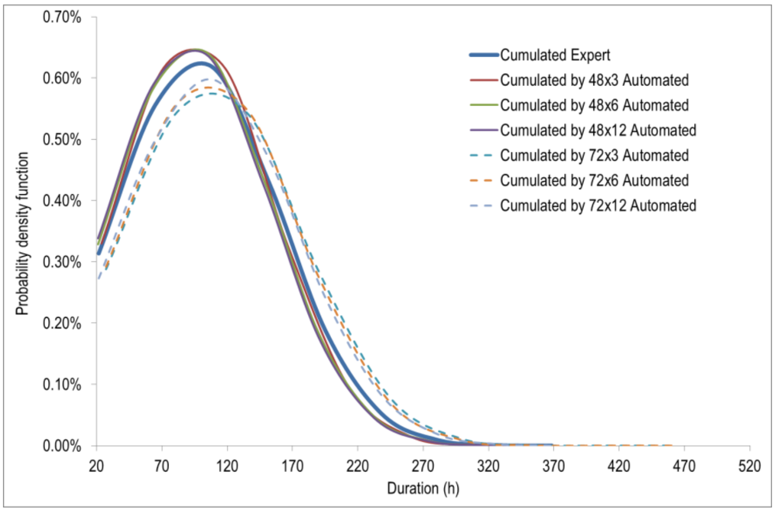

Fig. 4 Best fitting distribution models for the six samples of cumulated values calculated by the automated method and the sample by the expert method

From Table 4 the following two pieces of information can be drawn: 1) all E samples follow the log-normal distribution at high P-values. These values are the highest among all tested models; 2) all D samples show the highest P value for the Weibull distribution model. Figures 4 and 5 show the probability density functions of the best fitting models for the E and D samples. As it is evident, the probability density functions of the expert samples are much better mimicked by the 48 group of automated samples.

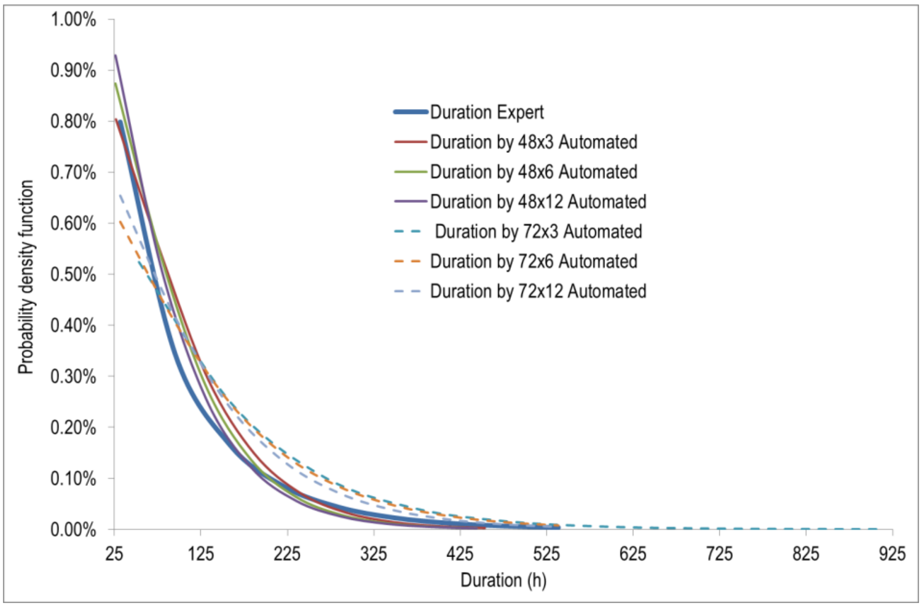

Fig. 5 Best fitting distribution models for the six samples of duration values calculated by the automated method and the sample by the expert method

Table 5 summarizes the results from statistical comparisons of the pairs of D and E samples from the expert and the automated methods. They aim at checking three null hypotheses: 1) Equal mean values (Fisher test is used), 2) Equal variances (ANOVA table and F test), 3) Population of the two samples (Kolmogorov-Smirnov's test). Thus, the null hypotheses are verified if no statistically meaningful differences between samples from the two procedures will be found at the 95% confidence level. Comparisons between E samples verify the three null hypotheses, but this is not true for D samples. The reason for this result is related to the Duration sample distributions: the Weibull distributions are not fully defined by the mean values and the variances, but rather by the two variables k and λ according to the following expression: $k/\lambda^k \;\; x^{(k-1)} \, e^{(-(x/\lambda)^k)}$. Thus, the first two null hypotheses can be applied only to Normal or LogNormal distributions. Nonetheless, the third null can be applied because the Kolmogorov-Smirnov test does not need the sample to follow a Gaussian distribution (being a non-parametric test). This null hypothesis is verified by D samples of expert and by the 48x12 combination of the automated procedure. Thus, these two samples belong to the same population based on the statistical tests. This also implies that they can be employed indifferently for tracing the rainfall empirical threshold lines. Moreover, this result highlights that the automated procedure is able to mimic the expert judgment by means of the combination device. This is true in the current study and for the 300 landslide events analyzed.

Table 5 Statistical comparisons of E and D samples: expert samples versus the 6 combinations drawn by the automated method. The null hypotheses are checked at 95% confidence

Pairs of independent samples	**No differences between Mean Values (Fisher test of Least Significance Difference)**	**No differences between Standard deviations (F test through ANOVA Table)**	**No differences in cumulative distribution functions (Kolmogorov-Smirnov test)**
Cumulated (expert vs 48x3)	0.74	0.33	0.99
Cumulated (expert vs 48x6)	0.46	0.51	0.93
Cumulated (expert vs 48x12)	0.3	0.51	0.65
Cumulated (expert vs 72x3)	0.05	0.25	0.06
Cumulated (expert vs 72x6)	0.05	0.42	0.08
Cumulated (expert vs 72x12)	0.1	0.39	0.21
Duration (expert vs 48x3)	0.5	Not applicable	0.02
Duration (expert vs 48x6)	0.7	Not applicable	0.04
Duration (expert vs 48x12)	0.24	Not applicable	0.3
Duration (expert vs 72x3)	0	Not applicable	0
Duration (expert vs 72x6)	0	Not applicable	0
Duration (expert vs 72x12)	0.0002	Not applicable	0.0003

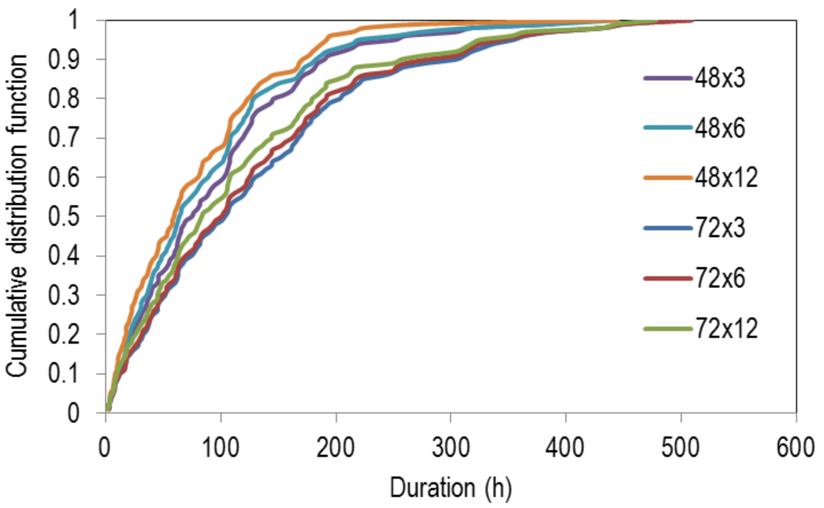

Fig. 6 Cumulative distribution function of duration samples of 300 values calculated through 6 combinations by the automated method

5) Results and discussion

According to the statistical analysis, only one D-E pair from the automated method is representative for the expert sample. In detail (Table 4), all the E samples from the automated method are representative for the expert one, but this is not true for D samples. The outcome is statistically evident but not at first glance for a non-expert user that calculated several combinations of ΔT and W. As previously noted, both the automated and the expert method select E values and derives D values. In doing this, D samples from the automated procedure are not all similar to the expert one. Working on the Italian territory, only the combination 48x12 seems to replace the D-E samples from the expert method. Thus, in order to select the expert-like D sample in advance, the inspection of D sample cumulative distribution functions CdF can be undertaken. Figure 6 shows the six CdF of D samples: the combinations 48x3, 48x6 and 48x12 have the most vertical S shape. Thus, the variance associated to the sample decreases when the CdF gets vertical. The 48x12 CdF shows to be the most vertical (Fig. 6), as it can be analytical appreciated through a coefficient of verticality C_V:

$$C_V = \frac{Q_{100}}{Q_{50}} \tag{1}$$

where Q are quantile values: Q_{100} is the fourth quartile and Q_{50} the second quartile. This index represents an additional measure of Duration sample dispersion. In order to choose the best D sample it will show the highest verticality, which means the less dispersion. In fact, calculating the C_V values for the six D samples we have:

C_{V48x3}= 5.37; C_{V48x6}= 6.9; C_{V48x12}= 7.46; C_{V72x3}= 4.92; C_{V72x6}= 5.06; C_{V72x12}= 5.72

In the case study, the highest C_V is related to the combination that statistically resembles the expert D sample. Such a rule can be used when no expert duration samples are available.

Conclusions

This study used statistical tests to verify that an automated method can simulate the calculation of rainfall events (in terms of D and E values) responsible for shallow landslide initiation accomplished by an expert through a manual procedure. The statistical study on 300 shallow landslides occurred in Italy indicates that:

(1) E and D paired samples from the two methods do not belong to the same population;

(2) E and D independent samples show that at least one combination of W and ΔT parameters implemented into the coded procedure provides E and D samples similar to the expert method;

(3) D samples are much variable than the corresponding E samples;~~,~~ ~~thus~~ therefore, the combination of D,E samples ~~must~~ should be chosen based upon the D sample variance (a C_V verticality coefficient was introduced for this purpose).

These outcomes confirm that the automated procedure simulate the expert judgment notwithstanding the systematic bias discussed in section 4.1. Hence, coded procedures ~~can~~ appear to be useful to minimize the human errors and to enable -non-expert users to calculate the D-E pairs. In the Italian territory, this procedure ~~can~~ could be adopted also at the regional scale provided that a large number of D,E sample data is available.

References

Baum, R.L., Godt, J.W., 2010. Early warning of rainfall-induced shallow landslides and debris flows in USA. Landslides 7, 259-272.

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 Geophysical Research 117, F04006. doi:10.1029/2012JF002367.

Bolley, S., Oliaro, P., 1999. Analisi dei debris flows in alcuni bacini campione dell'Alta Val Susa. GEAM 1999, 69 – 74.

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.

Brunetti, M.T., Peruccacci, S., Antronico, L., Bartolini, D., Deganutti, A.M., Gariano, S.L., Iovine, G., Luciani, S., Luino, F., Melillo, M., Palladino, M.R., Parise, M., Rossi, M., Turconi, L., Vennari, C., Vessia, G., Viero, A., Guzzetti, F., 2015. Catalogue of Rainfall Events with Shallow Landslides and New Rainfall Thresholds in Italy. Springer Special Series: "Engineering Geology for Society and Territory"; Volume 2: "Landslide Processes", pp. 1575-1579. doi: 10.1007/978-3-319-09057-3_280. ISBN: 978-3-319-09056-6.

Caine, N., 1980. The rainfall intensity-duration control of shallow landslides and debris flows, Geograf. Annal. 62A, 23–27.

Ceriani, M., Lauzi, S., Padovan, N., 1994. Rainfall thresholds triggering debris-flows in the alpine area of Lombardia Region, central Alps – Italy, Proc. Man and Mountain, Ponte di Legno (Italy), 123–139.

Cui, Y.J., Tang, C.S., Tang, A.M., Ta, A.N., 2014. Investigation of soil desiccation cracking using an environmental chamber. Italian Geotechnical Journal, 1, 9-20.

Dahal, R.K., Hasegaw, S., 2008. Representative rainfall thresholds for landslides in the Nepal Himalaya. Geomorphology, 100(3–4), 429–443.

Gariano S.L, Iovine G.G.R., Brunetti M.T., Peruccacci S., Luciani S., Bartolini D., Palladino M.R., Vessia G., Viero A., Vennari C., Antronico L., Deganutti A.M., Luino F., Parise M., Terranova O.G., Guzzetti F. 2012.Populating a catalogue of rainfall events that triggered shallow landslides in Italy. Rend. Online Soc. Geol. It., 21, 396-398.

Innes, J.N., 1983. Lichenometric dating of debris-flow deposits in the Scottish Highlands. Earth Surface Processes Landforms 8, 579–588.

ISPRA, 2013. Dati del Sistema Informativo di Carta della Natura alla scala 1:250.000.

Longobardi, A., Khaertdinova, E., 2015. Relating soil moisture and air temperature to evapotranspiration fluxes during inter-storm periods at a Mediterranean experimental site. J Arid Land 7(1), 27-36.

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.

Papa, N., Medina, V., Ciervo, F., Bateman, A., 2013. Derivation of critical rainfall thresholds for shallow landslides as a tool for debris flow early warning systems. Hydrology Earth System Sciences 17(10), 4095-4107.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen–Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644. doi:10.5194/hess-11-1633-2007.

Peruccacci, S., Brunetti, M. T., Luciani, S., Vennari, C., Guzzetti, F., 2012. Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central Italy, Geomorphology 139-140, 79–90.

R Core Team, 2013. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rossi, M., Peruccacci, S., Brunetti, M.T., Marchesini, I., Luciani, S., Ardizzone, F., Balducci, V., Bianchi, C., Cardinali, M., Fiorucci, F., Mondini, A.C., Reichenbach, P., Salvati, P., Santangelo, M., Bartolini, D., Gariano, S.L., Palladino, M., Vessia, G., Viero, A., Antronico, L., Borselli, L., Deganutti, A.M., Iovine, G., Luino, F., Parise, M., Polemio, M., Guzzetti F., Tonelli G., 2012. SANF: National warning system for rainfall-induced landslides in Italy. In: Eberhardt E., Froese C., Turner A.K. & Lerouil S. (Eds.), Landslides and Engineered Slopes. Proc. 11th Int. Symp. Landslides, Banff (Canada), 2, 1895-1899.

Sandersen, F., Bakkehøi, S., Hestnes, E., Lied, K., 1996. The influence of meteorological factors on the initiation of debris flows, rockfalls, rockslides and rockmass stability. In: Senneset, K. (Ed.), Landslides. A.A, Balkema, Rotterdam, pp. 97–114.

Segoni S., Rossi G., Rosi A., Catani F., 2014. Landslides triggered by rainfall: a semi-automated procedure to define consistent intensity duration threshold. Computers and Geosciences, 63, 123-131.

StatPoint Technologies Inc., 2013. Statgraphics Centurion XVI

Tarquini, S., Isola, I., Favalli, M., Mazzarini, F., Bisson, M., Pareschi, M.T., Boschi, E., 2007. TINITALY/01: a new Triangular Irregular Network of Italy. Annals of Geophysics 50, 407-425.

Tarquini, S., Vinci, S., Favalli, M., Doumaz, F., Fornaciai, A., Nannipieri, L., 2012. Release of a 10-m-resolution DEM for the Italian territory: Comparison with global-coverage DEMs and anaglyph-mode exploration via the web, Computers and Geosciences 38, 168-170. doi: doi:10.1016/j.cageo.2011.04.018.

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, 2399-2408.

Wieczorek, G.F., Guzzetti, F., 1999. A review of rainfall thresholds for triggering landslides. Proc. EGS Plinius Conference, Maratea (Italy), pp. 407 – 414.

Wilson, R.C., Mark, R.K., Barbato, G.E., 1992. Operation of real time warning system for debris flows in the San Francisco Bay area, California. In: Shen, H.W., Wen, F. (eds.), Hydraulic Engineering '93. American Society of Civil Engineers, San Francisco, CA, pp. 1908 – 1913.

Zhuang J., Cui P., Wang G., Chen X., Iqbal J., Guo X., 2015. Rainfall thresholds for the occurrence of debris flows in the Jiangjia Gully, Yunnan Province, China, ENGEO 4075. doi: 10.1016/j.enggeo.2015.06.006.