Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides

Luca PICIULLO[(1)], Stefano Luigi GARIANO[(2,3)(*)], Massimo MELILLO[(2)], Maria Teresa BRUNETTI[(2)], Silvia PERUCCACCI[(2)], Fausto GUZZETTI[(2)], Michele CALVELLO[(1)]

(1) Department of Civil Engineering, University of Salerno, Italy

(2) CNR IRPI, via Madonna Alta 126, 06128, Perugia, Italy

(3) Department of Physics and Geology, University of Perugia, via A. Pascoli, 06123, Perugia, Italy

* Corresponding author: gariano@irpi.cnr.it, phone: +39 075 5014424.

Keywords: early warning system, shallow landslide, rainfall threshold, EDuMaP method, Campania.

Abstract

A process chain for the definition and the performance assessment of an operational regional warning model for rainfall-induced landslides, based on rainfall thresholds, is proposed and tested in a landslide prone area in the Campania region (Southern Italy). A database of 96 shallow landslides triggered by rainfall in the period 2003-2010 and rainfall data gathered from 58 rain gauges are used. First, a set of rainfall threshold equations are defined applying a well-known frequentist method to all the reconstructed rainfall conditions responsible for the documented landslides in the area of analysis. Several thresholds at different exceedance probabilities (percentiles) are evaluated and 9 different percentiles combinations are selected for the activation of 3 warning levels. Subsequently, for each combination, the issuing of warning levels is computed by comparing, over time, the measured rainfall with the pre-defined warning level thresholds. Finally, the optimal percentiles combination to be employed in the regional early warning system, i.e., the one providing the best model performance in terms of success and error indicators, is selected employing the "event, duration matrix, performance" (EDuMaP) method.

1. Introduction

The literature reports several studies on early warning systems for the prediction of rainfall-induced landslides. They can be employed at “local” or “regional” scale (ICG 2012; Thiebes et al. 2012; Calvello and Piciullo 2016). Local warning systems address individual landslides (e.g., Lollino et al. 2002; Blikra 2008; Iovine et al. 2010; Intrieri et al. 2012; Michoud et al. 2013; Thiebes et al. 2013; Manconi and Giordan, 2015), while regional warning systems deal with populations of landslides in a region (e.g., Alfieri et a. 2012; Martelloni et al. 2012; Rossi et al. 2012; Segoni et al. 2014, 2015; Calvello et al. 2015a,b; Rosi et al., 2015; Stahili et al. 2015).

Regional landslide early warning systems are used to assess the probability of occurrence of rainfall-induced landslides over large areas, typically through the prediction and monitoring of meteorological variables, in order to warn authorities, civil protection personnel and the population. They can be schematized distinguishing among warning models and warning management strategies (Calvello and Piciullo, 2016). A regional landslide early warning model (ReLWaM) includes a regional correlation law (ReCoL) and a decision algorithm. A ReCoL is defined as a functional relationship between rainfall and landslides that can lead to the definition of rainfall thresholds for possible landslide occurrence (Guzzetti et al. 2007). A decisional algorithm contains a set of assumptions for defining the number of warning levels and of procedures linking rainfall thresholds to warning levels. ReCoL and warning models refer to the technical sphere of a regional landslide early warning system (Calvello et al., 2015b), whereas warning management considers aspects oriented to the social sphere, i.e. warning dissemination, communication strategy and emergency plan. Once the procedures to define and operate the ReLWaM are defined, a periodic analysis of the performance and an update of the ReCoL (i.e., of the rainfall thresholds, Rosi et al., 2015) are needed to improve the performance of the system and its reliability.

The evaluation of the performance of a ReLWaM is based on 2×2 contingency tables computed for the joint frequency distribution of observed and predicted landslides (e.g., Giannecchini et al., 2012; Martelloni et al. 2012; Peres and Cancelliere 2014; Staley et al. 2013; Lagomarsino et al. 2015; Greco et al. 2013; Segoni et al. 2014; Gariano et al. 2015; Rosi et al., 2015; Stähli et al. 2015). Segoni et al. (2015), Lagomarsino et al.

(2015) and Gariano et al. (2015) have proposed similar approaches to evaluate the reliability of rainfall thresholds for the prediction of rainfall-induced landslides, using back-analyses, contingency tables, and skill scores. However, in these cases, the model performance is assessed neglecting some important aspects which are peculiar to ReLWaM, among which (Calvello and Piciullo 2016): (i) the possible occurrence of multiple landslides in a warning area, (ii) the duration of the warning, (iii) the level of the warning in relation to the landslide spatial density in the warning area, and (iv) the relative importance that the system managers attribute to different types of errors (e.g., false positives and false negatives). Recently, Sättele et al. (2015; 2016) have proposed a framework for the evaluation of the effectiveness of an early warning system for all kinds of natural hazards. The framework starts from the assessment of the technical and the inherent reliability of the system, evaluated differently for automated and non-automated systems, and leads to an effectiveness analysis.

The main topic covered by this paper is how to employ rainfall thresholds into a reliable ReLWaM. To this aim, several questions need to be answered, such as: (i) which rainfall thresholds should be used in the landslide early warning system? (ii) How the thresholds should be selected? (iii) What is the optimal number of warning levels? (iv) To which warning level should correspond a rainfall threshold?

In an attempt to answer these questions, we propose a method based on a process chain in order to realize an objective procedure for the definition and the evaluation of a reliable threshold-based operational early warning system. First, we adopt a consolidated approach (Brunetti et al. 2010; Peruccacci et al. 2012; Gariano et al. 2015; Melillo et al. 2015, 2016) to define and validate empirical, cumulated event rainfall – rainfall duration (*ED*) thresholds for possible landslide occurrences. Afterwards, we propose a methodology for issuing warning levels, as a result of the comparison between measured rainfall and established thresholds. Finally, we assess the performance of the ReLWaM employing the EDuMaP (“Event, Duration Matrix, Performance”) method, proposed by Calvello and Piciullo (2016). We test the process chain into an area of 1619 km^2 in the Campania region (Southern Italy).

2. A process chain

2.1 Warning model: from rainfall thresholds to warning levels

The technical procedures of a reliable ReLWaM necessary to define and issue a certain warning level (WL) can be schematically resumed into five steps (Figure 1). Step 1 consists of defining and validating a set of rainfall thresholds with different exceedance probabilities (Section 2.2). Step 2 refers to the selection of rainfall thresholds for the activation of increasing warning levels in the ReLWaM. The higher is the warning level, the larger is the probability of landslide occurrence. In step 3, cumulated rainfall on different time intervals are calculated starting from rainfall measurements and compared with the rainfall thresholds associated to pre-identified warning levels, to issue the appropriate warning level, in step 4. Finally, in step 5, an evaluation of the ReLWaM performance (Section 2.3) in order to increase the reliability of the model through a periodical update of the warning levels is strictly necessary.

Figure 2 shows a hypothetical application of the procedure for issuing a WL. Hyetographs in the figure show the measured hourly rainfall, while inset graphs display three *ED* thresholds (black lines) that identify four increasing WL. The areas in green, yellow, orange, and red represent the combinations of cumulated rainfall, *E*, and duration, *D*, which belong to each WL. Starting from the time t=0 (starting evaluation time), the cumulated rainfall *E* is calculated for fixed antecedent intervals: 6 (Fig. 2a), 12 (Fig. 2b), 24, 36, and 48 hours (Fig. 2c). The resulting *E-D* condition of each antecedent interval (blue dot) belongs to a certain WL area in the inset graph. The maximum WL reached is emitted in the next 6 h (e.g., orange in Fig. 2d). The procedure is repeated after 6 h, with a new reference time t=0 (Fig. 2e). Again, the highest WL resulting from the antecedent rainfall conditions is issued for the following 6 hours (e.g., red in Fig. 2f). The procedure applied herein is based, as shown in figure 2d-f, on 6-hour-long steps, thus allowing an evaluation of the landslide WL four times per day.

2.2 Definition and validation of empirical rainfall thresholds

Empirical rainfall thresholds for possible landslide occurrence are defined through statistical analyses of past rainfall events that have resulted in landslides in a given study area. To obtain reliable thresholds, a large number of landslides for which the location and the time (or period) of the failure is known, and sufficiently accurate information on the rainfall responsible for landslide is needed. The geographical position and the time of the failure are usually affected by uncertainty. Thus, a class of geographical and of temporal accuracy is assigned to each landslide. Adopting a consolidated approach, three classes of geographic accuracy, G, are adopted, where the accuracy depends on the type and quality of the available information (Gariano et al. 2012). The first class (G_1) is attributed to landslides mapped with a geographic accuracy of 1 km^2, or less. The second (G_{10}) and the third (G_{100}) classes are attributed to landslides that are located with an accuracy of less than 10 km^2, and less than 100 km^2, respectively. Moreover, three classes of temporal accuracy, T, are defined. The first class (T_1) includes landslides for which the exact time of occurrence is known. The second and the third classes include landslides for which the part of the day (T_2) or the day of occurrence (T_3) were inferred.

To reconstruct the rainfall conditions responsible for landslides, the procedure proposed by Melillo et al. (2015, 2016) is applied to hourly rainfall measurements. First, starting from a rainfall record, and considering a minimum dry period (i.e., a period without rainfall, or with a negligible amount of rainfall) between two consecutive rainfall periods, all the rainfall events are singled out. A minimum dry period of 96 hours to distinguish the rainfall events in the wet season (from November to March), and of 48 hours to separate the rainfall events in the dry season (from April to October) are considered (Melillo et al. 2016). The rainfall conditions responsible for each landslide are automatically calculated considering the rainfall record from a representative rain gauge located in a buffer of 12 km from the landslide location. Criteria to select the rain gauge include: proximity, the elevation difference between the rain gauge and the landslide, and the local morphological setting (Melillo et al. 2015). Thus, the procedure calculates one or more rainfall conditions responsible for each landslide in the catalogue.

To determine the empirical rainfall thresholds, the frequentist method proposed by Brunetti et al. (2010), and modified by Peruccacci et al. (2012), is applied to all the reconstructed rainfall conditions responsible for the landslides. In this approach, the threshold curve is a power law equation linking the cumulated event rainfall E (in mm) to the rainfall duration D (in h),

$$E = (\alpha \pm \Delta\alpha) \cdot D^{(\gamma \pm \Delta\gamma)} \quad \text{(eq. 1)}$$

where, α is a scaling constant (the intercept), γ is the shape parameter (defining the slope of the power law curve), and $\Delta\alpha$ and $\Delta\gamma$ are the uncertainties of α and γ calculated using a "bootstrap" non-parametric statistical technique. The uncertainties associated to the thresholds depend on the number and the distribution of the empirical data points, and decrease as the number of the empirical data increase in the data set (Vennari et al. 2014).

To validate the thresholds, the method proposed by Gariano et al. (2015), that exploits a contingency table (Wilks 1995), Receiver Operating Characteristic (ROC) analysis (Fawcett 2006), and the related skill scores is adopted. In the contingency table, a "true positive" (TP) is an empirical (*D*,*E*) pair located above the threshold that has resulted in (at least) one landslide, and a "true negative" (TN) is an empirical (*D*,*E*) point below the threshold that has not resulted in known landslides. "False positives" (FP) occur when the (*D*,*E*) rainfall conditions exceeded the threshold and landslides did not occur (or where not reported). A "False negative" (FN) occurs when the (*D*,*E*) rainfall conditions were below the threshold and landslides occurred. The four contingencies are affected by biases caused by the lack of information on rainfall and/or landslide data (Gariano et al. 2015). Using the total number of TP, TN, FP, and FN, four skill scores are calculated, namely: (i) the Probability of Detection score, $POD = \frac{TP}{TP+FN}$, (ii) the Probability of False Detection score, $POFD = \frac{FP}{FP+TN}$, (iii) the Probability of False Alarm score, $POFA = \frac{FP}{TP+FP}$, and the Hansen and Kuipers (1965) skill score, $HK = \frac{TP}{TP+FN} - \frac{FP}{FP+TN} = POD - POFD$. The ROC analysis is usually performed constructing a ROC plot that shows the Probability of Detection, POD against the Probability of False Detection, POFD. In the ROC plane, the ROC curve is obtained varying the exceedance probability of the rainfall threshold and each point represents the prediction capability of the single threshold. For each rainfall threshold, the Euclidean distance δ between the point representing the threshold on the ROC curve and the upper left corner of the ROC plot – known as "perfect classification" (neither FN nor FP) – is calculated. The shorter the distance δ, the more suitable is the threshold, and consequently the model prediction skill.

To establish the optimal, “best performing” threshold that maximizes and/or minimizes the skill scores, the index Λ is defined as a linear combination of HK, POFA, and δ:

$$\Lambda = \lambda_1 \cdot HK - \lambda_2 \cdot POFA - \lambda_3 \cdot \delta \qquad \text{(eq. 2)}$$

where λ_1, λ_2, and λ_3 are positive scalar coefficients representing the weights of the individual skill scores, and $\lambda_1 + \lambda_2 + \lambda_3 = 1$. The combination of the skill scores that maximizes Λ represents the optimal compromise between the minimization of incorrect landslide predictions, and the maximization of the correct predictions.

2.3 Performance evaluation of the warning model

To evaluate the performance of the ReLWaM adopted within a regional landslides early warning system, the “Event, Duration Matrix, Performance” (EDuMaP) method proposed by Calvello and Piciullo (2016) is applied. EDuMaP comprises the following three successive steps:

- the analysis of the landslide (LE) and warning (WE) events;
- the definition and computation of a “duration matrix” that lists the time intervals associated with the occurrence of landslide events in relation to the emission of warning events;
- the evaluation of the performance of the early warning model using an established set of performance indicators.

In the first step, the values of the following 10 input parameters need to be specified:

1) number of WL used by the model;
2) thresholds used to differentiate among the *k* classes of LE on the basis of their spatial characteristics, $L_{\mathrm{den(k)}}$;
3) value of the time interval between the sending out of the first WL identified within a WE and the assumed beginning of the WE, t_{LEAD};
4) type of landslides addressed by the warning model, L_{typ};
5) time quantifying the maximum temporal gap among landslides included within a single LE, Δt_{LE};
6) time interval between the last landslide identified within a LE and the assumed

ending of the LE, t_{OVER};

7) area of analysis for which both landslides and warnings data are available, A;
8) subdivision of the area of analysis in m classes on the basis of the spatial criteria adopted to issue the warnings, ΔA;
9) temporal length of the databases for which both landslides and warnings data are available, ΔT;
10) minimum unit of time used to identify LE and WE, Δt (see also Appendix).

Thus, the assessment of the model performance requires the preliminary identification of the LE and WE from analyses carried out on the landslides and warnings databases. A LE is retrieved from the landslides database according to data, classification, spatial and temporal characteristics of the landslide records. In particular, a LE is obtained by grouping the collected landslides as a function of (Calvello and Piciullo, 2016): the landslide types (L_{typ}), the minimum interval between successive landslide events (Δt_{LE}), the temporal discretization of the analysis (Δt) and the over time (t_{OVER}) (Appendix). A WE is defined as a set of warning levels issued within a given warning zone (ΔA), grouped considering their temporal characteristics (Δt).

Concerning the second step, the number of rows and columns in the duration matrix is equal to the number of classes and levels defined for the LE and the WE, respectively. Figure 3 portrays a 4×4 duration matrix related to four levels of WE (no warning, WL_0; moderate warning, WL_1; high warning, WL_2; very high warning, WL_3) and four classes of LE (no landslides, no; few landslides, small event, S; several landslides, intermediate event, I; many landslides, large event, L). Each element d_{ij} of the duration matrix is computed, within the time frame of the analysis ΔT, as follows:

$$\mathrm{d}_{ij} = \sum_{\Delta T} (t_{ij}) \qquad \text{(eq. 3)}$$

where: i identifies the WE level, j identifies the LE class, t_{ij} is the time during which a WE of level i is concurrent with a LE of class j.

In the final step, two performance criteria (Fig. 4) are applied to assign a meaning to the elements of the duration matrix and to carry out the performance analysis. The "alert classification" criterion (Fig. 4a), employs an alert classification scheme derived from a standard contingency table, and it identifies correct predictions (CP), false alerts (FA), missed alerts (MA), and true negatives (TN). The issuing of the two highest levels of warning (WL_2 and WL_3) concurrently with the occurrence of the greatest classes of LE (I and L) are assumed as CP of the ReLWaM. The same for the issuing of the two lowest levels of warning (WL_0 and WL_1) and the simultaneous occurrence of the smallest classes of LE (no and S). FA and MA are incorrect prediction of the system and TN represent the absence of both warning and landslide occurrences. The "grade of accuracy" criterion (Fig. 4b) assigns a colour code to the components of the duration matrix in relation to the agreement between a given WE and a given LE. For instance, if the maximum WL is issued (i.e., WL_3) and only few landslides occur (i.e., the LE class is S), this should be considered a significant error of the warning model. Using this criterion, the elements are classified in four color-coded classes, as follows: green (Gre) for the elements which are assumed to be representative of the best model response, yellow (Yel) for elements representative of minor model errors, red (Red) for elements representative of a significant model error, and purple (Pur) for elements representative of a severe model error. Starting from the two performance criteria, several performance indicators can be derived (Calvello and Piciullo 2016). Table 1 lists the indicators considered in this work.

3. Case study

3.1 Landslide early warning system in Campania

The Campania region extends for 13,671 km^2 in southern Italy. The southern Apennines mountain range dominates the orography, exceeding 2000 m of elevation. A hilly landscape characterizes the eastern side of the region, whereas large plains separating isolated limestone and volcanic reliefs are present in the western part of the region. In the region, the mean annual rainfall ranges from 1000 to 2000 mm (Longobardi et al. 2016). Due to the rugged orography, severe storms are frequent in the region and result in abundant flash floods, debris flows, and shallow landslides (Cascini et al. 2008; Vennari et al. 2016, and references therein) that cause casualties and serious damage to urban areas and infrastructures. In the 50-year period 1950-2014, 286 persons were killed or went missing, 406 were injured, and more than 23,000 people were evacuated due to landslides in the region (http://polaris.irpi.cnr.it).

In Campania, a regional landslide early warning system exists as a part of the regional warning system developed and managed by the regional civil protection agency to deal with "hydraulic and geo-hydrological risks" (DPGR 299/2005). The system includes two phases: wheatear forecast and environmental monitoring. The first phase consists in issuing warnings based on numerical rainfall forecasts. For the purpose, the Campania region is subdivided into eight alert zones (AZ, Fig. 5) for weather forecast and early warning purposes, according to homogeneity criteria, which consider the following factors: hydrography, morphology, rainfall, geology, land-use, hydraulic and hydrogeological events, and administrative boundaries. The monitoring phase includes: (i) the evaluation of meteorological and hydrological events, (ii) the hydrological and weather forecast at steps of 6 hours, through now-casting techniques and rainfall-runoff modelling using real time parameters. The rainfall monitoring network encompasses 154 rain gauges and a meteorological radar.

3.2 **The test area**

Our test area is the *Camp-3* AZ (marked with number 3 in Fig. 5), extending for 1619 km^2 and encompassing 109 municipalities, 58 rain gauges. It includes the Lattari mountains, the Pizzo d'Alvano massif and the Picentini mountains (Fig. 5). Due to the presence of pyroclastic soil deposits mantling the carbonatic bedrock, the area is highly susceptible to rainfall-induced shallow landslides and debris flows (Di Crescenzo and Santo 2005; Cascini et al. 2008; Terranova et al. 2015; Napolitano et al. 2015). Indeed, it suffered some of the most catastrophic rainfall-induced landslide events in Europe. The most damaging events occurred on 25 October 1954 and caused, in the area of the *Sorrentino-Amalfitana* peninsula, 482 casualties, including 318 deaths, and more than 12,000 evacuees (http://polaris.irpi.cnr.it). The most recent catastrophic event is dated 4-5 May 1998. In those days, more than one hundred slope failures occurred over the slopes of the Pizzo d'Alvano massif and about two million m^3 of material were mobilized, causing 159 deaths, more than 6400 evacuees, and €500-million damage to buildings and infrastructure (Cascini 2004).

3.3 Catalogue of landslides

A catalogue of 305 rainfall-induced shallow landslides was compiled between January 2003 and December 2013 (11-years period) for the Campania region. Information on landslide occurrences were gathered from newspapers, internet and technical reports provided by local Fire Brigades and Civil Protection agency. The Authors are aware that additional landslides may have occurred in the area in the analysed time frame, although they may have not been reported, thus they are not included in the catalogue, due to lack of information.

Regarding the *Camp-3* alert zone, 140 rainfall-induced landslides were collected in the period from 2003 to 2013. The landslides archived in the catalogue occurred almost exclusively from September to March (129 out of 140), and were most abundant in January (32 landslides). The years with greatest number of recorded landslides (25) were 2009 and 2010. More than half of the landslides in the area (82 out of 140, 59%) were localized with a high geographic accuracy (G_1), 53 failures (38%) with a medium accuracy (G_{10}), and only 5 landslides (3%) with a low accuracy (G_{100}). The exact time of occurrence (T_1) is known for 86 landslides (61% of the total), on the contrary was inferred (T_2) for 32 landslides (23%). For the remaining 22 landslides (16%) in the catalogue, only the day of occurrence is known (T_3). Information about typology is not available for about half of landslides in the catalogue (67 out of 140). The remaining landslides were classified as rock falls (38), earth flows (12), debris flows (13) and (10) mudflows (sensu Hungr et al. 2014). The catalogue of 140 rainfall-induced landslides occurred in the *Camp-3* AZ was divided into two subsets: (i) a calibration set, listing 96 landslides occurred between January 2003 and December 2010, used to define the rainfall thresholds, and (ii) a validation set, listing 44 landslides occurred between January 2011 and December 2013, used to validate the thresholds.

4. Results and discussion

4.1 Rainfall thresholds

Adopting the procedure presented in Section 2.2 and using information on 96 landslides occurred in the *Camp-3* AZ between January 2003 and December 2010 (calibration set) and rainfall data recorded by 58 rain gauges, empirical rainfall thresholds for several exceedance probabilities (percentiles) were determined. Following Melillo et al. (2015, 2016), 201 multiple (*D*,*E*) rainfall combinations responsible for the 96 documented landslides were reconstructed. Figure 6 shows, in log-log coordinates, the 201 multiple combinations (blue points, calibration set) and the related rainfall thresholds at 1% ($T_{1,AZ3}$) and 5% ($T_{5,AZ3}$) exceedance probability levels. Multiple combinations associated to the landslides cover the range of duration $1 \leq D \leq 650$ h, which is the range of validity for the threshold, and the range of cumulated rainfall $5.6 \leq E \leq 249.5$ mm. Threshold parameters α, γ, $\Delta\alpha$, and $\Delta\gamma$ (eq. 1) for different exceedance probabilities (from 1% to 90%) are reported in Table 2. The table lists also the parameters for the thresholds at 5% exceedance probability level ($T_{5,Cam}$, α=10.1, γ=0.25, $\Delta\alpha$=1.1, $\Delta\gamma$=0.02) calculated using the dataset for the whole Campania region: 627 multiple conditions responsible for 305 landslides in the period 2003-2013. This threshold is reported in order to make a comparison with thresholds defined for other regions in southern Italy for similar periods (Calabria, Vennari et al. 2014; Sicily, Gariano et al., 2015). In particular, the $T_{5,Cam}$ is very similar to the one defined for Sicily for the period 2002-2011, whose parameters are: α=10.4, γ=0.27, $\Delta\alpha$=1.4, $\Delta\gamma$=0.03. On the other hand, $T_{5,Cam}$ is steeper (i.e., is characterized by a lower value of the γ parameter) than the threshold defined for Calabria for the period 1996-2011, whose parameters are: α=8.6, γ=0.41, $\Delta\alpha$=1.1, $\Delta\gamma$=0.03.

The thresholds defined for the *Camp-3* AZ were validated using 43 triggering rainfall conditions (red points in Fig. 7, validation set) responsible for 44 landslides occurred in the area between January 2011 and December 2013. Two landslides were associated to the same rainfall condition. For validation purpose, only one rainfall condition is associated with each landslide for the calculation of the values in the contingency table, as made by Gariano et al. (2015). The 43 rainfall conditions are in the range of duration $2 \leq D \leq 274$ h, and in the range of cumulated rainfall $17.4 \leq E \leq 142.6$ mm. In addition, 3995 rainfall events were reconstructed in the same period (green points in Fig. 7). These rainfall events are in the ranges of $1 \leq D \leq 274$ h and $1.2 \leq E \leq 190.2$ mm.

Table 3 summarizes the four contingencies (TP, FP, FN, TN) and the four skill scores (TPR, FPR, FAR, HK) for 10 thresholds, at different exceedance probabilities or percentiles (from 1% to 90%). The largest values for the HK, δ, and Λ indices were obtained by $T_{10,AZ3}$, that can be considered the optimal threshold, representing the best compromise between the minimum number of incorrect landslide predictions (FP, FN) and the maximum number of correct predictions (TP, TN).

4.2 From rainfall thresholds to warning levels

The methodology proposed in Section 2.1 was applied to the case study of *Camp-3* AZ for the period 2003-2013. Nine combinations of thresholds at different exceedance probabilities (Table 3) were considered for the issuing of WL, as reported in Figure 8. The first warning level (WL_0) can be defined by *ED* conditions not exceeding the lowest threshold. Then, *ED* conditions included between the first and the second threshold activate the second warning level (WL_1). Consequently, *ED* conditions, exceeding the second threshold and remaining below the third one, activate the WL_2. Finally, *ED* conditions exceeding the third threshold determine the issuing of the highest warning level (WL_3).

Starting from the 1 January 2003, at 00:00, considering steps of six hours, the antecedent rainfall conditions at time intervals of 6, 12, 24, 36 and 48 hours, for each rain gauge of the *Camp-3* AZ, were evaluated. The values obtained were compared with the percentiles combinations associated with the four WL. The highest WL threshold exceeded in at least one rain gauge defined the WL to be issued for the following 6-hour period to the entire *Camp-3* AZ. The procedure was employed at 6-hour steps for the whole period of the analysis, obtaining nine different sets of warnings, each set related to each combination of percentiles considered.

Table 4 lists the hours of activations per WL for each combination of percentiles in the period 2003-2013. As expected, the higher the percentile employed for a single WL, lower the number of hours of alert (defined as the hours of WL_1, WL_2, and WL_3). Evidently, raising the percentile associated to WL_i, with $i \neq 0$, and keeping the others unchanged, a decrease of hours for WL_i and an increase of WL_{i-1} is obtained.

4.3 Landslide and warning events analysis

As described in Section 2.3, the definition of a set of ten parameters is necessary to carry on the first step of the EDuMaP method, i.e. the events analysis. The parameters used to define and characterize the LE were kept constant for all the considered percentiles combinations. The recorded landslides were grouped into LE considering all rainfall-induced landslide, a minimum interval between successive landslide events Δt_{LE} = 24 h, a temporal discretization for the analysis Δt = 1 h and no over time (t_{OVER} = 0). Taking into account these parameters, 89 landslide events were defined in the *Camp-3* AZ (parameter A and ΔA) in the period 2003-2010 (parameter ΔT), derived by the 140 landslides collected in the catalogue. Table 5 lists the number of reconstructed LE per number of landslides. Most of the LE (62) report only one landslide (i.e., preceded and followed by 24 hours without landslides). The highest number of landslides composing a LE is seven. LE were grouped into four classes, based on the number of landslides belonging to each event (L_{den}). LE with up to two landslides were classified as small events (S). LE with a number of landslides between 3 and 9 were classified as intermediate events (I), and LE having more than nine landslides were considered large events (L). In the considered period, 75 LE were classified as small, 14 LE as intermediate, and no LE was classified as large (Table 6). Regarding the WE, nine different datasets were obtained from the nine combinations of the percentiles (Table 4), which produced a different duration matrix and consequently different values of performance indicators. For all the combinations the lead time (t_{LEAD}=0) was always set to zero.

4.4 Performance evaluation with EDuMaP

In order to define the optimal percentile combination to be employed as WL in a reliable ReLWaM, i.e., the combination that provides the best ReLWaM performance, the EDuMaP method was applied as last step of the process chain proposed (Step 5 of Fig. 1).

Table 7 and Figure 9 show the results obtained for the nine percentile combinations considering the element of the duration matrix in terms of "alert classification" (Fig. 4a), and "grade of accuracy" (Fig. 4b) criteria. The pairs $P_{3,10,50}$-$P_{1,10,50}$ and $P_{3,35,50}$-$P_{1,35,50}$ differ for the percentile used as threshold for WL_1. In terms of hours (Table 7) this affects CP and TN ("alert classification" criterion), and Yel and Gre ("grade of accuracy" criterion), due to the way the elements of the duration matrix were defined for each criteria. Higher values of CP and Yel were obtained for lower percentiles considered as WL_1 (e.g., comparing $P_{3,10,50}$ to $P_{1,10,50}$ and $P_{3,35,50}$ to $P_{1,35,50}$, Table 7, Fig.

9). This behaviour is due to a relocation of t_{ij} durations (see eq. 3) from the first to the second row of the matrix. The combinations $P_{1,10,50}$-$P_{1,35,50}$, $P_{3,10,50}$-$P_{3,35,50}$ and $P_{1,50,90}$-$P_{1,65,90}$-$P_{1,80,90}$, have different thresholds for WL_2 (Fig. 8). An increase of the threshold considered as WL_2 resulted, in terms of hours, in: a reduction of FA and Red, an increase of CP and Yel, and a slight variation of MA and Gre (Table 7, Fig. 9). The combinations $P_{1,50,65}$-$P_{1,50,90}$ and $P_{1,65,80}$-$P_{1,65,90}$ differ for the percentile considered as WL_3 (Fig. 8). An increase of WL_3 threshold implied a slight variation in terms of hours for CP, FA, Gre and Yel. On the contrary, a substantial difference, of one order of magnitude, is obtained for Red and Pur errors, with a reduction of the number of hours for severe model error.

The evaluation of performance indicators was conducted neglecting the element d_{11} of the duration matrix, that represents the number of hours without either landslides or warnings. Typically, the value of this element is orders of magnitude higher than the other elements of the matrix because it also includes all the hours without rainfall, for which a ReLWaM is not designed to deal with, specifically. Thus, d_{11} element is neglected in our analysis in order to avoid an overestimation of the performance. Table 8 and Figures 10-11 show the results in terms of performance indicators for the nine different percentiles combinations. Success (Fig. 10) and error (Fig. 11) performance indicators are plotted separately. Concerning the success indicators and in particular the efficiency index (I_{eff}), raising the percentile of WL_2 a general increase is observed, as it is evident comparing the pairs: $P_{3,10,50}$-$P_{3,35,50}$, $P_{1,10,50}$-$P_{1,35,50}$, and $P_{1,50,90}$-$P_{1,65,90}$-$P_{1,80,90}$ (Fig. 8). In particular, a 25% increment in the percentile related to the activation of the second WL, passing from $P_{3,10,50}$ to $P_{3,35,50}$ or from $P_{1,10,50}$ to $P_{1,35,50}$, corresponds to about 35% of increase of I_{eff}. Raising the percentile of 15%, from $P_{1,50,90}$ to $P_{1,65,90}$ and from $P_{1,65,90}$ to $P_{1,80,90}$, the I_{eff} shows an increment of about 5% (Fig. 10, Table 8). The percentile of WL_3 do not influence the I_{eff} (i.e. $P_{1,50,65}$- $P_{1,50,90}$ and $P_{1,65,80}$- $P_{1,65,90}$) because CP and FA are subjected to a very small variation and the MA value is orders of magnitude lower than the firsts. Regarding WL_1, if its percentile is reduced, a positive effect can be observed on the I_{eff}, because CP increase.

The hit rate (HR_L) is very high for all the percentile combinations (Fig. 12), slightly lower than 100% (Table 8), due to a minor number of hours of MA compared to those of CP (Table 7). The positive predictive power (PP_w) shows variations similar to the I_{eff}, because they just differ in the calculation for the MA, which in this case, are very low compared to the other elements of the duration matrix (Fig. 10). The odds ratio (OR), which can be considered as a rate between correct and predictions, increases as a function of the reduction of FA and MA, and the increment of CP (Fig. 11).

Among the error indicators, the missed alert rate (R_{MA}) and the false alert rate (R_{FA}) are dependent respectively by the hours of MA and FA. The first are very low, probably also dependent by the low number of LE of class intermediate and large. The FA substantially decrease, in terms of hours, as the percentiles increase (Table 8 and Fig. 11). The error rate (ER) and the probability of serious mistakes (P_{SM}) are evaluated excluding the element d_{11} in order to exclusively evaluate the errors due to the functioning of the system, avoiding underestimation. For this case study, these indicators are principally dependent by the value assumed by FA, which show high value of Red and Pur errors for low percentile combinations of WL_2 and WL_3 (Figs. 9 and 11). It is important to point out, that for our case study, in the period of analysis, $d_{4,1}$ is the only contribution to Pur errors, because components $d_{1,4}$ and $d_{2,4}$ are null for all the nine percentile combinations (Table 7) since there are not LE classified as large.

Among the nine combinations of percentiles, $P_{1,80,90}$ provide the best results in terms of both success and error performance indicators (Table 8). However, the performance analysis was conducted with a database of landslide of an 11-years period, during which large LE and few intermediate LE did not occurr. Thus, the performance analysis was oriented, basically, on defining the percentiles combinations with both low FA and high CP. The aim is obtained raising the percentiles for WL_2 and WL_3 (i.e., reducing FA) and decreasing the percentile of WL_1 (i.e., increasing CP).

5. Conclusions

As a general goal, this paper focuses on the definition of an operational and reliable regional landslide early warning system, ReLWaM. "Operational" in terms of considering all the assumptions and procedures needed to technically operate a regional early warning model, including: (i) the definition of rainfall thresholds and warning levels, (ii) the evaluation of monitored rainfall, (iii) the comparison between rainfall and thresholds, (iv) the production and issuing of warnings. "Reliable" since it is based on an optimal definition of warning level thresholds, resulting in the best early warning performance. To deal with these issues, a process chain in 5 steps, for the definition and the performance assessment of an operational regional warning system for rainfall-induced landslides, based on rainfall thresholds, is proposed. The methodology defined in Section 2 can be used to issue a certain level of warning at 6-hour steps, by comparing the monitored rainfall with warning level thresholds. The highest threshold exceeded defines the warning level to be issued for the following 6 hours in a certain warning zone. The paper and the proposed procedure do not address the following important warning management issues: risk perception, policy adopted to communicate with the people at risk, evacuation procedures, monitoring network and instruments used to issue the warnings.

As a specific target, an operational and reliable ReLWaM for rainfall-induced landslides was conceived for the *Camp-3* AZ, in the Campania region, Southern Italy, through the application of the process chain herein proposed. Empirical rainfall thresholds at different exceedance probabilities (percentiles) were defined applying a well-known frequentist method, and validated using ROC analysis and skill scores. The optimal threshold (i.e., the one with exceedance probability equal to 10%) defined with the ROC analysis cannot employed by itself in a ReLWaM, due to the high probability of false alerts. For this reason, nine percentiles combinations were separately considered as thresholds for the activation of 3 warning levels. Each percentiles combination resulted in a distinct WE database. Finally, in order to define the optimal percentiles combination, i.e., the one that provides the best ReLWaM performance, the EDuMaP method was applied. The performance analysis carried out for different percentile combinations highlight a high influence of percentile related to the activation of WL_2 on the I_{eff} index. The OR index probably represents the most effective indicator to describe the results, as it relates correct prediction and incorrect ones. The percentile combination $P_{1,80,90}$ resulted the best solution for the ReLWaM employed in the *Camp-3* AZ, because it yields good results both in terms of success and error performance indicators and the highest value of OR. It is worth highlighting that the database, for the period of analysis, has a low number of intermediate LE and no large LE, thus, the choice of the best performance was principally oriented on the FA reduction and CP increment. The high number of hours of FA can be justified by the way the WLs are defined in Section 2.1. In fact, a certain WL is issued if the related threshold is exceeded in at least one rain gauge in the area of analysis. This approach can be considered conservative, as it leads to a high number of FA but results in fewer MA. Moreover, in the analyses herein proposed, only the monitored rainfall was compared with warning level thresholds. More generally, other variables could be considered as relevant for triggering landslides and could be taken into account in the process of landslide forecasting and performance analysis. The best percentiles combination obtained represents the optimal solution for the database available at the time the performance analysis was carried out. Therefore, a continuous collection of data, an update of the thresholds and a periodic performance assessment are necessary to maintain a high reliability of the ReLWaM.

In conclusion, our work provides the following important insights:

- the definition of a set of rainfall thresholds at different exceedance probabilities (percentiles) is a fundamental issue;
- a decisional algorithm is needed for passing from rainfall thresholds to WL to be issued in a certain warning zone;
- a percentile combination, without a performance evaluation, is not sufficient to obtain a reliable and performative ReLWaM;
- the definition of a single threshold is not the most reliable solution to be employed in ReLWaM;
- the performance evaluation revealed the importance of OR in selecting the optimal combination of percentiles to be employed as warning levels in a ReLWaM;
- a performance evaluation is strictly connected to the availability of a landslide catalogue and to the accuracy of the information included in it.

Acknowledgements

This work was financially supported by the PhD programme of the civil engineering department of the University of Salerno, for LP, and by a grant from the Italian National Department for Civil Protection, for SLG and MM. G. Iovine and O. Terranova (CNR

IRPI) contributed to find information on landslide occurrences. We are grateful to the Regional Functional Centre of Civil Protection of Campania for providing rainfall data, and to the fire brigades of Avellino, Benevento, Caserta, Napoli, and Salerno for providing information on landslide occurrences. We thank two anonymous reviewers for their criticisms and comments that have helped us to improve the paper.

References

Alfieri L, Salamon P, Pappenberger F, Wetterhalll F, Thielen J (2012) Operational early warning systems for water-related hazards in Europe. Environ Sci Policy 15(1):35-49. doi:10.1016/j.envsci.2012.01.008

Blikra LH (2008) The Åknes rockslide: monitoring, threshold values and early-warning. In: Chen Z et al (eds) Landslides and Engineered Slopes, edited by: Taylor & Francis Group, London, pp 1089–1094. doi:10.1201/9780203885284-c143

Brunetti MT, 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. doi:10.5194/nhess-10-447-2010

Calcaterra D, Parise M, Palma B (2003). Combining historical and geological data for the assessment of the landslide hazard: a case study from Campania, Italy. Nat Hazards Earth Syst Sci 3(1/2):3-16. doi:10.5194/nhess-3-3-2003

Calvello M, d'Orsi RN, Piciullo L, Paes NM, Magalhaes MA, Coelho R, Lacerda WA (2015a) The community-based alert and alarm system for rainfall induced landslides in Rio de Janeiro, Brazil. In: Engineering Geology for Society and Territory "Landslide Processes", Proceedings of the XII Int. IAEG Congress, Turin, Italy 2:653–657. doi:10.1007/978-3-319-09057-3_109

Calvello M, d'Orsi RN, Piciullo L, Paes N, Magalhaes MA, Lacerda WA (2015b) The Rio de Janeiro early warning system for rainfall-induced landslides: analysis of performance for the years 2010–2013. Int. J. Disast Risk Reduc 12:3–15. doi:10.1016/j.ijdrr.2014.10.005

Calvello M, Piciullo L (2016) Assessing the performance of regional landslide early warning models: the EDuMaP method. Nat Hazards Earth Syst Sci 16:103-122. doi:10.5194/nhess-16-103-2016

Cascini L (2004) The flowslides of May 1998 in the Campania region, Italy: the scientific emergency management. Ital Geotech J 2:11–44. ISSN:0557-1405

Cascini L, Cuomo S, Guida D (2008) Typical source areas of May 1998 flow-like mass movements in the Campania region, Southern Italy. Eng Geol 96:107–125. doi:10.1016/j.enggeo.2007.10.003

Di Crescenzo G, Santo A (2005) Debris slides-rapid earth flows in the carbonate massifs of the Campania region (Southern Italy): Morphological and morphometric data for evaluating triggering susceptibility. Geomorphology 66(1-4):255-276. doi:10.1016/j.geomorph.2004.09.015

DPGR n 299 del 30/06/2005, Decreto del Presidente della Giunta Regionale della Campania: Il Sistema di Allertamento Regionale per il rischio Idrogeologico e Idraulico ai fini di protezione civile, Bollettino ufficiale della regione Campania, n. speciale 01/08/2005, Campania, Italy.

Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27:861-874. doi:10.1016/j.patrec.2005.10.010

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. doi:10.5194/nhess-12-829-2012

Gariano SL, Brunetti MT, Iovine G, Melillo M, Peruccacci S, Terranova OG, Vennari C, Guzzetti F (2015) Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, Southern Italy. Geomorphology 228:653-665. doi:10.1007/s11069-014-1129-0

Gariano SL, Iovine G, Brunetti MT, Peruccacci S, Luciani S, Bartolini D, Palladino MR, Vessia G, Viero A, Vennari C, Antronico L, Deganutti AM, Luino F, Parise M, Terranova O, Guzzetti F, (2012) Populating a catalogue of rainfall events that triggered shallow landslides in Italy. Rend Online Soc Geol It 21:396-398. ISSN 2035-8008

Greco R, Giorgio M, Capparelli G, Versace P (2013) Early warning of rainfall-induced landslides based on empirical mobility function predictor. Eng Geol 153:68–79. doi:10.1016/j.enggeo.2012.11.009

Guzzetti F, Peruccacci S, Rossi M, Stark CP (2007) Rainfall thresholds for the initiation of landslides in Central and Southern Europe. Meteorog Atmos Phys 98:239–267. doi:10.1007/s00703-007-0262-7

Hanssen AW, Kuipers WJA (1965) On the relationship between the frequency of rain and various meteorological parameters. Koninklijk Nederlands Meteorologisch Institut Meded Verhand 81: 2-15.

Hungr O, Leroueil S, Picarelli L (2014) The Varnes classification of landslide types, an update. Landslides 11, 2, 167-194. doi:10.1007/s10346-013-0436-y

International Centre of Geohazards: Guidelines for landslide monitoring and early warning systems in Europe – design and required technology (2012) In: Project SafeLand "Living With Landslide Risk in Europe: Assessment, Effects of Global Change, and Risk Management Strategies", D4.8, http://www.safeland-fp7.eu (last access: April 2015)

Intrieri E, Gigli G, Mugnai F, Fanti R, Casagli N (2012) Design and implementation of a landslide early warning system. Eng Geol 147/148:124–136. doi:10.1016/j.enggeo.2012.07.017

Intrieri E, Gigli G, Casagli N, Nadim F (2013) Brief communication "Landslide Early Warning System: toolbox and general concepts". Nat Hazards Earth Syst Sci 13:85–90. doi:10.5194/nhess-13-85-2013

Iovine G, Lollino P, Gariano SL, Terranova O (2010) Coupling sensitivity limit equilibrium analyses and real-time monitoring to refine a landslide surveillance system in Calabria (Southern Italy). Nat Hazards Earth Syst Sci 10:2341–2354. doi:10.5194/nhess-10-2341-2010

Lagomarsino D, Segoni S, Rosi A, Rossi G, Battistini A, Catani F, Casagli N (2015) Quantitative comparison between two different methodologies to define rainfall thresholds for landslide forecasting. Nat Hazards Earth Syst Sci 15:2413–2423. doi:10.5194/nhess-15-2413-2015

Lollino G, Arattano M, Cuccureddu N (2002) The use of the automatic inclinometric system for landslide early warning: the case of Cabella Ligure (North-Western Italy). Phys Chem Earth 27:1545–1550. doi:10.1016/S1474-7065(02)00175-4

Longobardi A, Buttafuoco G, Caloiero T, Coscarelli R (2016) Spatial and temporal distribution of precipitation in a Mediterranean area (southern Italy). Environ Earth Sci 75(189). doi:10.1007/s12665-015-5045-8

Manconi A, Giordan D (2015) Landslide early warning based on failure forecast models: The example of the Mt. de la Saxe rockslide, northern Italy. Nat Hazards Earth Syst Sci, 15(7), 1639-1644. doi:10.5194/nhess-15-1639-2015

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

Melillo M, Brunetti MT, Peruccacci S, Gariano SL, 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. doi:10.1007/s10346-015-0630-1

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

Michoud C, Bazin S, Blikra LH, Derron M-H, Jaboyedoff M (2013) Experiences from site-specific landslide early warning systems. Nat Hazards Earth Syst Sci 13(2659–2673). doi:10.5194/nhess-13-2659-2013

Napolitano E, Fusco F, Baum RL, Godt JW, De Vita P (2015) Effect of antecedent-hydrological conditions on rainfall triggering of debris flows in ash-fall pyroclastic mantled slopes of Campania (southern Italy). Landslides. doi:10.1007/s10346-015-0647-5

Peres DJ, Cancelliere A (2014) Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach, Hydrol Earth Syst Sci, 18, 4913–4931, doi:10.5194/hess-18-4913-2014

Peruccacci S, Brunetti MT, 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. doi:10.1016/j.geomorph.2011.10.005

Rosi A, Lagomarsino D, Rossi G, Segoni S, Battistini A, Casagli N (2012) Updating EWS rainfall thresholds for the triggering of landslides. Nat Hazards 78: 297-308. doi:10.1007/s11069-015-1717-7.

Rossi M, Peruccacci S, Brunetti MT, Marchesini I, Luciani S, Ardizzone F, Balducci V, Bianchi C, Cardinali M, Fiorucci F, Mondini AC, Reichenbach P, Salvati P, Santangelo M, Bartolini D, Gariano SL, Palladino M, Vessia G, Viero A, Antronico L, Borselli L, Deganutti AM, Iovine G, Luino F, Parise M, Polemio M, Guzzetti F (2012) SANF: National warning system for rainfall-induced landslides in Italy. In: Eberhardt E, Froese C, Turner AK, Leroueil S (Eds.) Landslides and Engineered Slopes: Protecting Society through Improved Understanding. Taylor & Francis Group, London, 1895-1899. ISBN 978-0-415-62123-6.

Sättele M, Bründla M, Straubb D (2015) Reliability and effectiveness of early warning systems for natural hazards: Concept and application to debris flow warning. Rel Eng Syst Safety, 142, 192-202. doi:10.1016/j.ress.2015.05.003

Sättele M, Bründl M, Straub D (2016) Quantifying the effectiveness of early warning systems for natural hazards. Nat Hazards Earth Syst Sci, 16:149–166. doi: 10.5194/nhess-16-149-2016

Segoni S, Battistini A, Rossi G, Rosi A, Lagomarsino D, Catani F, Moretti S, Casagli N (2015) Technical Note: An operational landslide early warning system at regional scale based on space–time-variable rainfall thresholds. Nat Hazards Earth Syst Sci 15:853–861. doi:10.5194/nhess-15-853-2015.

Segoni S, Rosi A, Rossi G, Catani F, Casagni N (2014) Analysing the relationship between rainfalls and landslides to define a mosaic of triggering thresholds for regional scale warning systems. Nat Hazards Earth Syst Sci 14: 2637-2648. doi:10.5194/nhess-14-2637-2014.

Stähli M, Sättele M, Huggel C, McArdell BW, Lehmann P, Van Herwijnen A, Berne A, Schleiss M, Ferrari A, Kos A, Or D, Springman SM (2015) Monitoring and prediction in early warning systems for rapid mass movements. Nat Hazards Earth Syst Sci 15:905–917. doi:10.5194/nhess-15-905-2015, 2015

Staley DM, Kean JW, Cannon SH, Schmidt KM, Laber JL (2012) Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California. Landslides 10:547-562. doi:10.1007/s10346-012-0341-9

Terranova OG, Gariano SL, Iaquinta P, Iovine G (2015) ^{GA}SAKe: forecasting landslide activations by a genetic-algorithms-based hydrological model. Geosci Model Develop 8:1955-1978. doi:10.5194/gmd-8-1955-2015

Thiebes B, Bell R, Glade T, Jäger S, Mayer J, Anderson M, Holcombe L (2013) Integration of a limit-equilibrium model into a landslide early warning system. Landslides 11:859–875. doi:10.1007/s10346-013-0416-2

Thiebes B, Glade T, Bell R (2012) Landslide analysis and integrative early warning-local and regional case studies. In Eberhardt E (Eds.) Landslides and Engineered Slopes: Protecting Society through Improved Understanding. Taylor & Francis Group, London, 1915–1921.

Vennari C, Gariano SL, Antronico L, Brunetti MT, 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. doi:10.5194/nhess-14-317-2014

Vennari C, Parise M, Santangelo N, Santo A (2016) First evaluation of the damage related to alluvial events in torrential catchments of Campania (southern Italy), based on a historical database. Nat Hazards Earth Syst Sci Discuss. doi:10.5194/nhess-2015-355

Wilks DS (1995) Statistical Methods in the Atmospheric Sciences. Academic Press, 467 pp.

Table captions

Table 1. Performance indicators used for the analysis. Criteria a) and b) refer to Fig. 4. Modified from Calvello and Piciullo (2016).

Table 2. Parameters α and γ of *ED* thresholds at the 5% exceedance probability level, and related uncertainties Δα and Δγ determined using different subsets of data.

Table 3. Contingencies (TP, FP, FN, TN) and skill scores (POD, POFD, POFA, HK, δ, Λ) calculated for thresholds at different exceedance probabilities; best scores are shown in bold.

Table 4. Number of hours (and related percentages) of activation for the four warning levels per each combination of threshold percentiles.

Table 5. Number of landslide events (LE) as a function of the number of landslides.

Table 6. Number of landslide events (LE) pertaining to the four classes.

Table 7. Values in hours of the duration matrix elements in terms of “alert classification” (criterion a, Fig. 4a) and the “grade of accuracy” (criterion b, Fig. 4b) criteria. Best values are shown in bold.

Table 8. Performance indicators obtained for each percentile combination. Best values are shown in bold.

Figure captions

Figure 1. Steps of the procedure proposed to define and to issue warning levels within a regional early warning model for rainfall-induced landslides.

Figure 2. Example of the algorithm adopted to determine warning levels (WL) using rainfall thresholds. (a) Cumulated rainfall (blue bars) for the antecedent period of 6 h before the evaluation time t=0 and related WL (blue dot in the inset graph). (b) Cumulated rainfall (blue bars) for the antecedent periods of 6 and 12 h before the evaluation time t=0 and related WL (blue dots in the inset graph). (c) Cumulated rainfall (blue bars) for the antecedent periods of 6, 12, 24, 36, and 48 h before the evaluation time t=0 and related WL (blue dots in the inset graph). (d) Cumulated rainfall (blue bars) for the antecedent periods of 6, 12, 24, 36, and 48 h before the new evaluation time t=0, 6 h later, and related WL (blue dots in the inset graph).

Figure 3. Structure of the duration matrix with four levels of WE (key: no, no warning; M, moderate warning; H, high warning; VH, very high warning) and four classes of LE (key: no, no landslides; few landslides, small event, S; several landslides, intermediate event, I; many landslides, large event, L). Modified from Calvello and Piciullo (2016).

Figure 4. Alert classification (a) and grade of accuracy (b) performance criteria used for the analysis of the duration matrix with four classes of WE (key: no, no warning; M, moderate warning; H, high warning; VH, very high warning) and four classes of LE (key: no, no landslides; few landslides, small event, S; several landslides, intermediate event, I; many landslides, large event, L). Modified from Calvello and Piciullo (2016).

Figure 5. Map of the *Camp-3* Alert Zone (*Sorrentino-Amalfitana* peninsula, Pizzo d'Alvano massif, Picentini mountains) showing shaded relief, classes of altitude (m a.s.l.), 58 rain gauges used in this study (blue triangles) and 140 rainfall-induced landslides (red circles). The main toponyms are also indicated. The insets show the location of Campania region in Italy and the subdivision of the region into eight alert zones for civil protection purposes.

Figure 6. Multiple *ED* rainfall conditions (multiple combinations) responsible for 96 landslides in the *Camp-3* AZ and related rainfall thresholds at 1% ($T_{1,AZ3}$) and 5%

($T_{5,AZ3}$) exceedance probability levels. Shaded areas portray uncertainty associated with the threshold curves. Data are in log-log coordinates.

Figure 7. Rainfall duration vs. cumulated event rainfall conditions in *Camp-3* AZ in the period 2011-2013, compared with thresholds (blue solid lines) at 1%, 5%, 10%, 20%, and 50% exceedance probability levels (indicated by the numbers in the labels), determined using the calibration set. Red points are 43 *ED* conditions associated with the triggering of shallow landslides in the validation period. Green points are 3995 rainfall events for which information on triggered landslides is not available. Grey points are 159 rainfall events with duration exceeding the range of validity of the thresholds (D > 650 h). Data are in log-log coordinates.

Figure 8. Extents of the 4 warning levels (WL0, WL1, WL2, WL3) as a function of the nine combinations of percentiles (P) of thresholds.

Figure 9. Percentage of CA, MA, FA, and TN (criterion A) and of Pur, Red, Yel, and Gre (criterion B) obtained for the nine considered percentile combinations.

Figure 10. Bar chart showing the values of success indicators for each percentile combination. Efficiency index (I_{eff}), hit rate (HR_L), predictive power (PP_W), and threat score (TS) values are shown as percentages (green bars). The absolute values for the odds ratio (OR) are also reported (brown bars, on secondary vertical axes in inverse order).

Figure 11. Bar chart showing the percentage values of error indicators for each percentile combination: misclassification rate (MR), missed alert rate (R_{MA}), false alert rate (R_{FA}), error rate (ER), and probability of serious errors (P_{SM}).

Appendix

Variables and acronyms used in text.

Acronym	Description
A	Area of analysis
AZ	Alert zone
CA	Correct Alert
D	Rainfall duration (h)
E	Cumulated event rainfall (mm)
EDuMaP	Event, Duration Matrix, Performance
ER	Error rate
FA	False Alert
FN	False Negative
FP	False Positive
G	Geographic accuracy
Gre	Green error
HK	Hansen and Kuipers skill score
HR_L	Hit Rate
I_{eff}	Efficiency Index
$L_{den(k)}$	Landslide density criterion
LE	Landslide event
L_{typ}	Landslide type
MA	Missed Alert
MR	Misclassification rate
OR	Odd Ratio
P	Percentile combination
POD	Probability of Detection score
POFA	Probability of False Alarm score
POFD	Probability of False Detection score
PP_W	Predictive Power
P_{SM}	Probability of Serious Mistakes
Pur	Purple error
RE	Rainfall event
ReCoL	Regional Correlation Law
Red	Red error
ReLWaM	Regional Landslide Warning Model
R_{FA}	False Alert Rate
R_{MA}	Missed Alert Rate
ROC	Receiver Operating Characteristic
T	Temporal accuracy
$T_{5,Cam}$	Rainfall threshold at 5% exceedance probability for the Campania region 2003-2013
$T_{1,AZ3}$	Rainfall threshold at 1% exceedance probability for the *Camp-3* alert zone 2003-2010
$T_{3,AZ3}$	Rainfall threshold at 3% exceedance probability for the *Camp-3* alert zone 2003-2010
$T_{5,AZ3}$	Rainfall threshold at 5% exceedance probability for the *Camp-3* alert zone 2003-2010
$T_{10,AZ3}$	Rainfall threshold at 10% exceedance probability for the *Camp-3* alert zone 2003-2010
$T_{20,AZ3}$	Rainfall threshold at 20% exceedance probability for the *Camp-3* alert zone

	2003-2010
$T_{35,AZ3}$	Rainfall threshold at 35% exceedance probability for the *Camp-3* alert zone 2003-2010
$T_{50,AZ3}$	Rainfall threshold at 50% exceedance probability for the *Camp-3* alert zone 2003-2010
$T_{65,AZ3}$	Rainfall threshold at 65% exceedance probability for the *Camp-3* alert zone 2003-2010
$T_{80,AZ3}$	Rainfall threshold at 80% exceedance probability for the *Camp-3* alert zone 2003-2010
$T_{90,AZ3}$	Rainfall threshold at 95% exceedance probability for the *Camp-3* alert zone 2003-2010
$T_{5,AZ3,03\text{-}11}$	Rainfall threshold at 5% exceedance probability for the *Camp-3* alert zone for the period 2003-2011
$T_{5,AZ3,03\text{-}12}$	Rainfall threshold at 5% exceedance probability for the *Camp-3* alert zone for the period 2003-2012
$T_{5,AZ3,03\text{-}13}$	Rainfall threshold at 5% exceedance probability for the *Camp-3* alert zone for the period 2003-2013
t_{LEAD}	Lead time
t_{OVER}	Over time
TN	True Negative
TP	True Positive
TS	Threat Score
WE	Warning Event
WL	Warning Level
Yel	Yellow error
α	Scaling parameter (intercept) of the rainfall threshold
γ	Shape parameter (slope) of the rainfall threshold
δ	Distance between the point representing a rainfall threshold and the upper left corner of the ROC plot
ΔA	Spatial discretization adopted for warnings
Δt_{LE}	Minimum interval between landslide events
Δt	Temporal discretization of analysis
ΔT	Time frame of analysis
λ	Scalar coefficient
Λ	Index representing a combination of HK, POFA and δ