

Landslides (2020) 17:61–77
 DOI 10.1007/s10346-019-01270-z
 Received: 25 April 2018
 Accepted: 27 August 2019
 Published online: 10 September 2019
 © Springer-Verlag GmbH Germany
 part of Springer Nature 2019

Hongjuan Yang · Fangqiang Wei · Zhenfeng Ma · Haiyan Guo · Pengcheng Su · Shaojie Zhang

Rainfall threshold for landslide activity in Dazhou, southwest China

Abstract Dazhou is one of the landslide-prone areas in southwest China. Since rainfall is the primary factor responsible for slope instability in this region, empirical rainfall thresholds may be effective for landslide warning. Thresholds using the intraday rainfall and the effective antecedent rainfall up to 7 days prior to landslide occurrence have been defined for this area. The study region was divided into 6 alert zones. For each zone, the daily rainfall and landslide data in the monsoon season during the period of 2003–2010 were analyzed, and the logistic regression method was employed to fit the dataset and provide diverse probability thresholds. Subsequently, the optimal threshold was identified by receiver operating characteristic (ROC) analysis. Furthermore, two classes of criteria and the associated indicators were used to determine the three rainfall thresholds for the warning model with four classes of alert (i.e., no warning, moderate warning, high warning, and very high warning). Finally, these thresholds were validated with dataset for the period 2011–2015. It suggests that the optimal threshold derived from ROC analysis, with the probability ranging from 4 to 10%, is suitable for issuing moderate warning, while the probability corresponding to the thresholds for high warning and very high warning varies in the ranges 30–55% and 75–95%, respectively. Performance of these thresholds in the period 2011–2015 is comparable to that in the period 2003–2010, indicating the robustness of them and the potential for practical use.

Keywords Rainfall threshold · Landslide · Logistic regression · ROC analysis

Introduction

Landslides extensively occur in the mountainous regions worldwide and have induced numerous casualties and property losses. In the areas susceptible to landslides, it is a cost-efficient tool to establish landslide early warning systems for hazard mitigation (Glade and Nadim 2014). Rainfall thresholds have been commonly employed in existing warning systems (Keefer et al. 1987; Aleotti 2004; Baum and Godt 2010; Segoni et al. 2015). This is mainly attributed to the significant roles of rainfall in facilitating slope instability: increasing the soil moisture, reducing the matric suction, and raising the pore water pressure (Chung et al. 2017). Additionally, the availability of weather forecast makes the prediction of potential landslide activity possible (Vallet et al. 2016).

Two types of rainfall threshold models have been used for landslide warning: physically based models and empirical models (Guzzetti et al. 2007). The first type of models are developed by simulating the hydrological process within the soil mass and analyzing the stability of the slopes (Wu et al. 2015; Kanjanakul et al. 2016). They can provide the specific time and location of potential landslides. However, it is difficult to get the hydrological and geotechnical parameters required to run these models on a regional scale. Thus, physically based models tend to be applied in smaller areas (Posner and Georgakakos 2015; Bezak et al. 2016).

Empirical models define the rainfall conditions for landslide initiation in a region by analyzing historical landslide events and the associated rainfall data. Without considering underlying conditions like geomorphology, land use, and lithology, these models cannot give the exact location of possible landslides. Nonetheless, they can be used in combination with landslide susceptibility zonation techniques to help the associated agencies to assess landslide risks (Segoni et al. 2015).

To define rainfall thresholds, the following parameters have been commonly employed to characterize the rainfall conditions: (i) rainfall intensity (I) (Guzzetti et al. 2008; Mathew et al. 2014; Segoni et al. 2014; Ma et al. 2015; Giannecchini et al. 2016; Rosi et al. 2016); (ii) rainfall duration (D), which is usually linked to I ; (iii) accumulated event rainfall (E) (Vennari et al. 2014; Gariano et al. 2015; Segoni et al. 2015; Lainas et al. 2016; Melillo et al. 2016; Rossi et al. 2017); and (iv) antecedent rainfall calculated for various time periods (Sepúlveda and Padilla 2008; Kanungo and Sharma 2014; Ma et al. 2014; Elias 2016). The I and E parameters can be normalized by mean annual precipitation or rainy-day-normal to concern the effect of climate conditions. The determination of which rainfall parameters are adopted is affected by the landslide and rainfall data available. For instance, the application of I and D requires rainfall measurements recorded at hourly or smaller time steps. Moreover, researchers might tend to use their preferable parameters. For example, some researchers adopt the I – D threshold, while others employ the E – D threshold, although these two types are interconvertible.

In general, it is impossible to divide the rainfall conditions into a 100% landslide field and a 100% non-landslide field. Therefore, two methods have been proposed to define rainfall thresholds (Lagomarsino et al. 2015). In the first method, the threshold is taken as the lower boundary of rainfall conditions that have triggered landslides (Bui et al. 2013; Lainas et al. 2016). Thus, it represents the minimum rainfall needed to initiate a landslide and is conservative in operational use for the great number of false alarms (Peres and Cancelliere 2014). The second approach identifies the threshold by a trade-off between the maximization of correct forecasts and the minimization of incorrect forecasts (Staley et al. 2013; Giannecchini et al. 2016). Statistical techniques, such as the Bayesian approach and the logistic regression analysis, are usually employed to derive thresholds for different exceedance probability levels or for diverse possibilities of landslide occurrence. The best-performing threshold is finally defined using contingency tables, receiver operating characteristic (ROC) analysis, and the related skill scores.

This study aims to define the rainfall threshold for landslide activity in Dazhou, which is a landslide-prone area in southwest China whereas little attention has been paid to the regional rainfall threshold. Concerning the relevant role of the antecedent rainfall in landslide triggering, we use this parameter in combination with the intraday rainfall to represent the rainfall condition. Initially, the rainfall and landslide data are systematically analyzed with

logistic regression. Then, we obtain the optimal threshold by ROC analysis. Furthermore, the thresholds for different warning levels are derived following a procedure for performance evaluation of models with multiple warning levels. Finally, the thresholds are validated with additional data from the study region.

Study area and materials

Environmental setting of the study area

Dazhou is located in the northeast of Sichuan Province, southwest China (Fig. 1), and extends for an area of 16,591 km². It consists of 7 county-level units: Wanyuan, Xuanhan, Dachuan, Tongchuan, Kaijiang, Qu, and Dazhu, with a total population of 5.6 million.

Lying in the zone where the Daba Mountain and the Sichuan Basin intersect, the terrain elevation of Dazhou decreases from 2399 m in the north to 250 m in the south (Fig. 1). The north part of the region is featured by mountainous landforms, where the relative relief usually goes above 1000 m. The middle and the south parts are dominated by low mountains and hills, with the relative relief varying from 700 to 1000 m and flatlands scattering in the valleys. In general, mountains are the primary landforms in the study area and account for 70.7%. Hills and flatlands account for 28.1% and 1.2%, respectively (Tang et al. 2010). Dazhou is located in the composite part of the Daba Mountain and the East Sichuan arc-like fold belts, where fractures and folds are well developed. Mudstone, shale, sandstone, and limestone formed in Triassic, Jurassic, or Cretaceous are distributed across the study area. Quaternary overburdens have been developed on some slopes from

mudstone and shale due to weathering, with the depth fluctuating between 3 and 9 m (Xiang 2015).

Climate in this area is typically subtropical, humid, and monsoon-affected. The mean annual temperature varies from 14.7 to 17.6 °C, with the south part warmer than the north. The mean annual precipitation changes between 1076 and 1270 mm, with the north slightly wetter than the south. The annual precipitation fluctuates greatly from 1 year to another. Moreover, precipitation concentrates in the monsoon season, namely, April–October. Geological disasters have widely occurred in this region due to the steep terrain, broken rock mass, and strong rainfall events during the monsoon season. A total of 1814 potential geohazard sites, which might bring damages to 102,828 inhabitants, have been identified (Liu et al. 2014).

Landslide and rainfall data

Landslide data used in this study were collected by the Sichuan Provincial Environmental Monitoring Center from 2003 to 2015. The database consists of the main information on a total of 1843 landslides: occurrence date, geographical location, landslide scale, and damage induced. Altogether, these landslides caused 28 deaths, 5 people missing, 66 injured, and a direct economic loss of 409 million RMB. Four landslides were eliminated in our research because the exact dates of their occurrences were not known. In recent years, the study area has undergone rapid industrialization and urbanization. Human activities such as roadbuilding and mining have increased the risk of landslide hazards. However, among the 372 landslides recorded with trigger

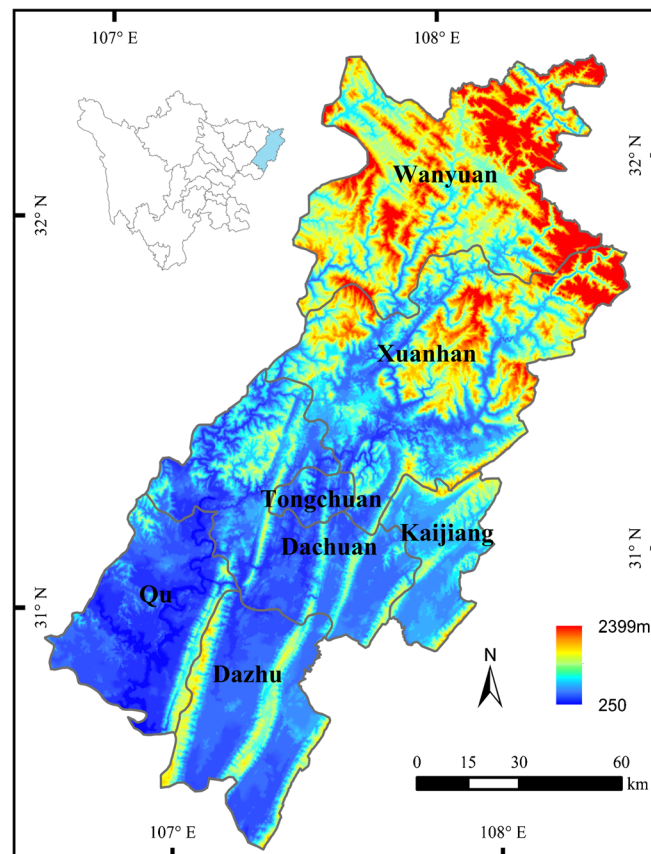


Fig. 1 The geographical location and terrain elevation of Dazhou. Names of the 7 county-level units are also shown

information, only 3 were induced by human activities and the others were triggered by rainfall. Considering the low proportion of human-induced landslides, it is assumed that all the remaining slope failures without recorded triggering factors were initiated by rainfall. Therefore, a total of 1836 landslides are available, with the locations shown in Fig. 2. In terms of landslide scale, the category of individual landslides rather than the exact volume is contained in the database. It shows that landslides pertaining to small-sized ($< 10^5 \text{ m}^3$), medium-sized ($10^5\text{--}10^6 \text{ m}^3$), large-sized ($10^6\text{--}10^7 \text{ m}^3$), and super-large-sized type ($> 10^7 \text{ m}^3$) account for 79.0%, 16.0%, 4.5%, and 0.5%, respectively. Distribution of each type of landslides in different administrative units is listed in Table 1. Landslide typology (e.g., shallow landslides, deep-seated landslides, etc.) is not listed in the database and thus was not considered in the present study.

Daily rainfall data measured by 6 rain gauges were employed for threshold analysis. Figure 2 depicts the locations of these gauges. The mean annual precipitation in the 2003–2015 period is 1294 mm, with the temporal distribution summarized in Fig. 3. Due to the high proportion of rainfall accumulated in the monsoon season (88.9%), almost all landslides were triggered in this period with only six exceptions. The number of landslides triggered in July is 1053 (accounts for 57.4%) and is much higher than the number for the other months. This is primarily attributed to the great amount of rainfall in July (accounts for 17.6%). Moreover, extreme rainstorms might lead to a soaring number of slope failures. An example can be found in mid-July 2010, when the recorded accumulated event rainfall was 449 mm in Wanyuan and 194 mm in

Xuanhan and it induced a total of 480 landslides in these two regions. In operational regional early warning systems, alerts are commonly issued based on administrative units. Concerning the distribution of rain gauges (Fig. 2), Tongchuan and Dachuan were combined into one alert zone (AZ), while each of the other 5 county-level administrative regions was taken as an independent alert zone. For simplicity, these alert zones are labeled AZ1 through AZ6, as illustrated in Fig. 2. The corresponding rain gauge situated in each alert zone was taken as the reference gauge.

The dataset was divided into two subsets: a calibration set including 1426 landslides triggered during the 2003–2010 period, and a validation set covering 410 landslides initiated over the 2011–2015 period. The effects of the 12 May 2008 Wenchuan earthquake should be mentioned. The earthquake has greatly impacted the rainfall thresholds for geohazards in the stricken zones thereafter (Yu et al. 2014). To evaluate the influence of the earthquake on rainfall threshold in Dazhou, spatial and temporal distribution of landslides before and after the earthquake was analyzed. On the one hand, the study area was divided into two parts by the equidistance line which is 250 km from the main fault of the earthquake (Fig. 2). A total of 978 landslides were reported during the 2008–2015 period, and 51.6% of them were located in the west part. This proportion is close to that during the period 2003–2007 (50.8%). On the other hand, inter-annual variation in precipitation and landslide occurrence was investigated (Fig. 3b). Considering that the numerous landslides in 2010 were mainly triggered by the extraordinary storm event in Wanyuan and Xuanhan detailed above, there is no evidence that the study area has become more

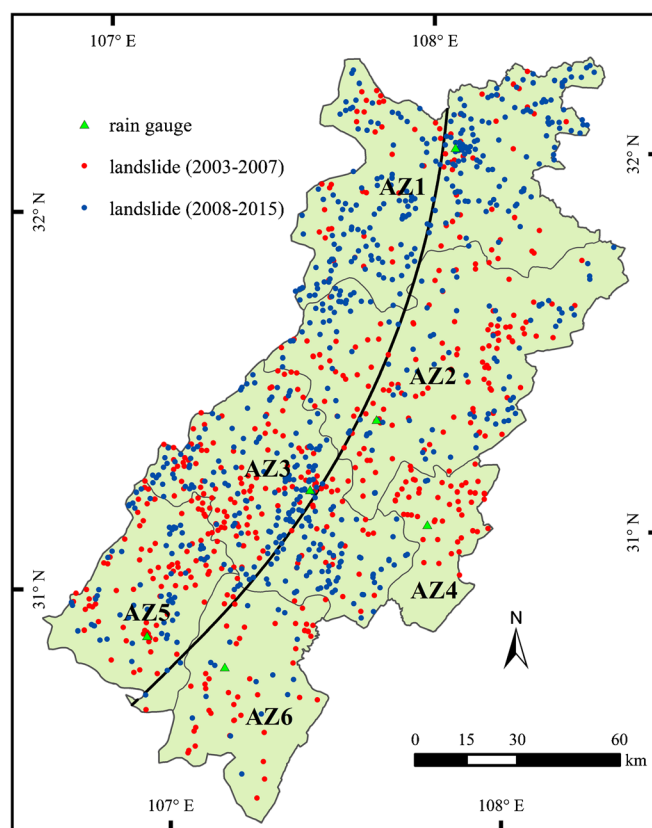


Fig. 2 Distribution of the landslides triggered during the period of 2003–2015 in Dazhou and locations of the rain gauges used in this study. Gray lines are boundaries of the 6 alert zones (AZ1–AZ6), while the bold black line is an equidistance line (250 km) from the main fault in the 12 May 2008 Wenchuan earthquake

Table 1 Landslide distribution in terms of their sizes in different administrative units

| Unit | Small | Medium | Large | Super-large | Total |
|-----------|-------|--------|-------|-------------|-------|
| Wanyuan | 532 | 70 | 10 | 4 | 616 |
| Xuanhan | 227 | 97 | 42 | 2 | 368 |
| Dachuan | 389 | 34 | 11 | 1 | 435 |
| Tongchuan | 58 | 5 | 3 | 2 | 68 |
| Kaijiang | 39 | 31 | 5 | 0 | 75 |
| Qu | 138 | 52 | 10 | 0 | 200 |
| Dazhu | 67 | 5 | 2 | 0 | 74 |

susceptible to landslides after the earthquake. In addition, no co-seismic landslides were reported in Dazhou for the lower seismic intensity (VI). Therefore, rainfall threshold for landslide occurrence in the study area was considered to be unchanged after the earthquake in this research.

Methodology

Rainfall parameters

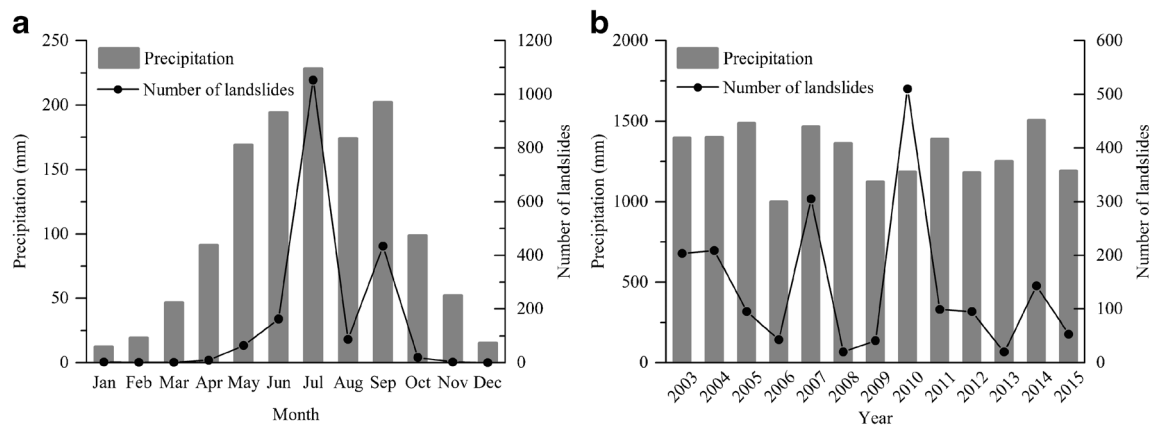
The antecedent rainfall can impact slope stability by changing soil moisture content and correspondingly affecting the geotechnical parameters. Thus, it was taken as one parameter for threshold definition, denoted by R_a . Another parameter is the intraday rainfall (R_o), which is usually the direct predisposing factor for slope failure. Moisture preserved in the soil mass will decline due to the evapotranspiration and discharge processes. Therefore, the effective antecedent rainfall rather than the accumulated antecedent rainfall was employed. Following Bruce and Clark (1966), it is written as:

$$R_a = \sum_{i=1}^n k^i R_i \quad (1)$$

where R_i is daily rainfall measured the i th day prior to landslide occurrence, n is the number of days considered, and k is the decay factor.

The value for the decay factor is affected by multiple factors, such as the evaporation capacity, the vegetation fraction, and the hydrological characteristics of the regolith. Different values ranging from 0.72 to 0.86 have been proposed by investigators, whereas the value 0.84 is used most extensively (Zhuang et al. 2014). Although this value was derived from the hydrological data measured in Ottawa (Bruce and Clark 1966), satisfactory performance was obtained when it was used in the rainfall threshold analysis for landslide warning in New Zealand (Crozier and Eyles 1980). We also employed $k = 0.84$ in the present study.

In terms of the number of days considered for computing R_a , values proposed in literature extend from 3 to 120, while values smaller than 30 are more commonly employed (Kanungo and Sharma 2014; Elias 2016). In some researches, R_a calculated for different time spans are plotted versus the intraday rainfall or the accumulated event rainfall, and the time span using which the rainfall conditions with landslide occurrence are more easily discriminated is finally adopted (Bui et al. 2013; Mathew et al. 2014). Therefore, we used different time periods varying between 3 and 30 days to compute R_a . We found that significant linear relationships exist among various R_a values. Figure 4 exhibits the slope and the determination coefficient when the R_a values computed for diverse periods are linearly correlated with that derived for 30 days. The slope is 0.90 when n is equal to 7, indicating that the corresponding rainfall is representative. It consists with the finding of Zhuang et al. (2014) in Xi'an, China, that the influence of

**Fig. 3** Distribution of precipitation and landslide occurrence during the year (a) and during the 2003–2015 period (b) in Dazhou

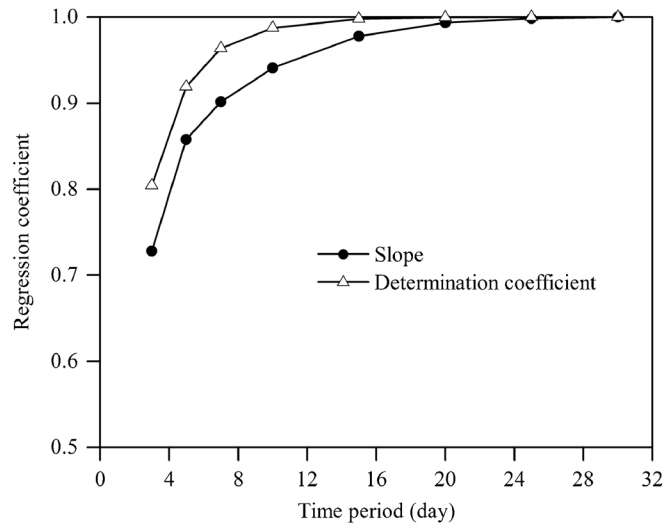


Fig. 4 The slope and determination coefficient when the effective antecedent rainfall computed for different periods is linearly correlated with that derived from a 30-day period

a rainfall event on soil moisture becomes negligible after 7 days. As a consequence, we used $n = 7$ in this research.

Logistic regression

Logistic regression has been widely used to discriminate rainfall conditions with landslide occurrence from those without slope failures (Glade et al. 2000; Chang et al. 2008; Frattini et al. 2009; Mathew et al. 2014; Giannecchini et al. 2016). In this approach, probability of landslide occurrence (p) is expressed as a sigmoid function of the linear combination of rainfall parameters:

$$\ln\left(\frac{p}{1-p}\right) = \alpha_1 R_o + \alpha_2 R_a + \alpha_o \quad (2)$$

where α_i is the regression coefficient.

For each alert zone, rainy days (with $R_o \geq 0.1$ mm) during April–October in the period of 2003–2010 were classified into two groups: landslide days (LDs) and non-landslide days (NLDs). Logistic regression was carried out using rainfall parameters associated with the two groups of days, and the coefficients in Eq. (2) were derived correspondingly.

ROC analysis

ROC analysis is an effective tool to evaluate the performance of classification models. It works on the basis of a contingency table, which expresses the four possible outcomes when a threshold model is applied. A true positive (TP) means that the rainfall condition exceeded the threshold and at least one landslide occurred. A false positive (FP) represents an instance that the rainfall condition exceeded the threshold but no landslides were recorded. An (R_a, R_o) data point located below the threshold is counted as true negative (TN) if no landslides were observed; otherwise, it is considered false negative (FN). In an early warning system, false positives and false negatives are false alarms and missed alarms, respectively.

Using the number of the four classes of outcomes, the following skill scores can be computed:

- Probability of detection (POD) measures the proportion of landslide days that are correctly forecasted: $POD = TP/(TP + FN)$;
- Probability of false detection ($POFD$) measures the proportion of non-landslide days which are erroneously predicted: $POFD = FP/(FP + TN)$;
- Probability of false alarm ($POFA$) is the ratio between the number of false alarms and the total number of forecasts: $POFA = FP/(TP + FP)$;
- Hanssen and Kuipers (1965) skill score (HK) evaluates the accuracy for the instances with and without landslides and is calculated as the difference between POD and $POFD$: $HK = POD - POFD$.

In the ROC space (Fig. 5), each data point represents the prediction capability of a rainfall threshold. Given thresholds for different probabilities of landslide occurrence, an ROC curve will be derived. It represents the forecasting ability of the logistic regression model, the performance of which can be evaluated using the area under curve (AUC). The diagonal line from the lower left corner to the upper right corner in the ROC graph represents the strategy of random guessing a class, and the corresponding AUC value is equal to 0.5. The upper left corner of the ROC graph (the black dot in Fig. 5) represents the best classification, i.e., $POD = 1$ and $POFD = 0$. The shorter the distance between the data point and the best classification, the better the performance of the corresponding threshold. To get the best-performing threshold, the index Λ defined by Gariano et al. (2015) was employed, which is expressed as a linear combination of HK , $POFA$, and the Euclidean distance δ from the perfect classification:

$$\Lambda = \lambda_1 \cdot HK - \lambda_2 \cdot POFA - \lambda_3 \cdot \delta \quad (3)$$

where λ_1 , λ_2 , and λ_3 are weight coefficients of individual skill scores, and $\lambda_1 + \lambda_2 + \lambda_3 = 1$. In our study, the same weight was

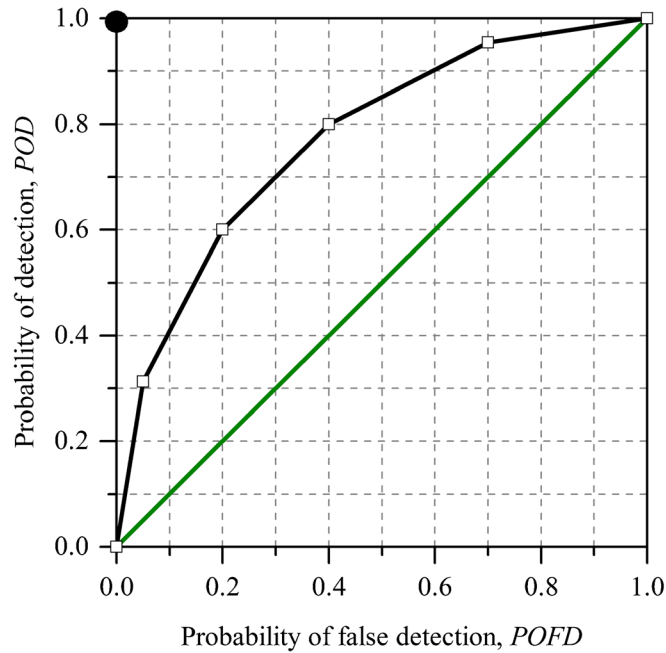


Fig. 5 ROC space with a hypothetical curve. The black dot represents the best classification, while the diagonal line (in green) represents the strategy of random guessing a class

given to the three skill scores, i.e., $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$. The combination of the scores that maximizes Λ represents the best compromise between maximizing correct predictions and minimizing false predictions.

Performance evaluation of the model with multiple warning levels

For a given region, not only the probability of landslide occurrence but also the magnitude of the possible hazard, in terms of the number, the scar area, the volume, or the density, increases when

| a Alet classification criterion | | Classes of landslide events | | | |
|--|-----------------|-----------------------------|-----------|-----------|-----------|
| | | no | S | I | L |
| Warning levels | WL ₀ | <i>TN</i> | <i>CP</i> | <i>MA</i> | <i>MA</i> |
| | WL ₁ | <i>CP</i> | <i>CP</i> | <i>MA</i> | <i>MA</i> |
| | WL ₃ | <i>FA</i> | <i>CP</i> | <i>CP</i> | <i>CP</i> |
| | WL ₃ | <i>FA</i> | <i>FA</i> | <i>CP</i> | <i>CP</i> |

| b Grade of accuracy criterion | | Classes of landslide events | | | |
|--------------------------------------|-----------------|-----------------------------|------------|------------|------------|
| | | no | S | I | L |
| Warning levels | WL ₀ | <i>Gre</i> | <i>Yel</i> | <i>Red</i> | <i>Pur</i> |
| | WL ₁ | <i>Yel</i> | <i>Gre</i> | <i>Red</i> | <i>Pur</i> |
| | WL ₃ | <i>Red</i> | <i>Yel</i> | <i>Gre</i> | <i>Yel</i> |
| | WL ₃ | <i>Pur</i> | <i>Red</i> | <i>Yel</i> | <i>Gre</i> |

Fig. 6 Alert classification (a) and grade of accuracy (b) criteria to assess the performance of a warning model with four classes of alert (Piciullo et al. 2017)

Table 2 Indicators used for evaluating the performance of a model with multiple levels of alert (Piciullo et al. 2017)

| Indicator | Symbol | Formula |
|---------------------------------|-----------|----------------------|
| Efficiency index | I_{eff} | $CP/(CP + MA + FA)$ |
| Hit rate | HR_L | $CP/(CP + MA)$ |
| Predictive power | PP_W | $CP/(CP + FA)$ |
| Probability of serious mistakes | P_{SM} | $Pur/(CP + MA + FA)$ |

the rainfall event becomes heavier (Gao et al. 2018). Consequently, a model with multiple warning levels considering landslide magnitude might be of greater value for landslide warning. In this study, both landslide number and the category of landslide size were used to evaluate the hazard magnitude and four classes of landslide events were specified following Piciullo et al. (2017): no landslides (no), small event (1–2 small-sized landslides, S), intermediate event (3–9 small-sized landslides or ≥ 1 medium-sized landslide, I), and large event (≥ 10 small and medium-sized landslides or ≥ 1 large and super-large-sized landslide, L). Correspondingly, four levels of warning may be issued: no warning (WL_0), moderate warning (WL_1), high warning (WL_2), and very high warning (WL_3), and three rainfall thresholds, denoted as RT_1 , RT_2 , and RT_3 , are required.

Calvello and Piciullo (2016) proposed a method to evaluate the performance of models providing multiple warning levels, which was further modified by Piciullo et al. (2017) and applied in Campania, Italy. It was also employed in the present research. In this method, two criteria are employed to evaluate the performance of the warning model. One is the “alert classification” criterion using a classification scheme developed from a two-by-two confusion matrix, as shown in Fig. 6a. Correct predictions (CP) are given to the situations that the two lowest levels of warning were issued and the two smallest landslide events occurred or the opposite situations. True negatives (TN) represent no warning and no landslide occurrence. Both false alerts (FA) and missed alerts (MA) are erroneous forecasts. The other criterion is the “grade of accuracy” criterion, which uses four color codes to represent the accuracy of the predictions: green (no error, *Gre*), yellow (minor error, *Yel*), red (significant error, *Red*), and purple (severe error, *Pur*), as depicted in Fig. 6b.

Using these criteria, Calvello and Piciullo (2016) defined 14 performance indicators. Eight of them have been adopted by Piciullo et al. (2017). Concerning the functional relationships

between different indicators, only four indicators used by Piciullo et al. (2017) were employed in our study, as listed in Table 2. Rainfall thresholds RT_1 , RT_2 , and RT_3 can be defined for the warning model based on these indicators.

Results and discussions

Results of logistic regression

Table 3 summarizes the number of landslide days (LDs) and non-landslide days (NLDs) for each alert zone (AZ) during the period 2003–2010. The landslide incidence is highest in AZ3 (10%) while it is lowest in AZ4–AZ6 (<5%). Rainfall conditions associated with the LDs and NLDs are plotted in the (R_o , R_a) space in Fig. 7, represented by the red and the green points, respectively. The values of R_o are mainly in the range $0 \leq R_o \leq 150$ mm, and the R_a values are mostly in the range $0 \leq R_a \leq 200$ mm. The red points and the green ones overlap in a large part of the graph. To be more specific, no landslides were recorded in some days with relatively large R_o or R_a values, while some landslides were triggered when the R_o and R_a values were relatively small. This is probably caused by the spatial heterogeneity in rainfall. Only one rain gauge, commonly located in the river valley, is available in each AZ. Rainfall in the mountainous and the hilly regions might be significantly different from that in the vicinity of the gauge. In other words, precipitation data employed are occasionally incapable to represent rainfall conditions at the landslide locations. Nevertheless, Fig. 7 exhibits the trend that landslide incidence increases with the increase in R_o and R_a , suggesting the potential of using these data for rainfall threshold determination.

Table 3 also lists the regression coefficients in Eq. (2), which were derived using the Statistical Product and Service Solutions software. Eq. (2) can be rewritten as:

$$R_o + \beta R_a = R(p) \quad (4)$$

$$\beta = \frac{\alpha_2}{\alpha_1}, R(p) = \frac{1}{\alpha_1} \left(\ln \left(\frac{p}{1-p} \right) - \alpha_0 \right)$$

which means that the iso-probability rainfall condition can be represented by a straight line with a slope of $-\beta$ and a y-intercept of $R(p)$ in the R_o – R_a space. Figure 7 depicts the R_o – R_a relationships for the 10%, 50%, and 90% probability levels. The iso-probability lines have the slopes $-0.83 \leq -\beta \leq -0.51$, indicating that the intraday rainfall prevails over the effective antecedent rainfall in landslide initiation. The value of β can reflect the

Table 3 Regression coefficients of the logistic regression model

| Alert zone | Number of LDs | Number of NLDs | α_0 | α_1 | α_2 | β | p_L | p_H |
|------------|---------------|----------------|------------|------------|------------|---------|-------|-------|
| AZ1 | 34 | 687 | – 5.131 | 0.031 | 0.025 | 0.804 | 0.6% | 89.0% |
| AZ2 | 59 | 686 | – 4.518 | 0.041 | 0.034 | 0.828 | 1.9% | 85.6% |
| AZ3 | 71 | 621 | – 3.875 | 0.043 | 0.026 | 0.595 | 2.6% | 94.8% |
| AZ4 | 23 | 711 | – 5.297 | 0.050 | 0.025 | 0.507 | 0.7% | 71.2% |
| AZ5 | 28 | 709 | – 4.874 | 0.032 | 0.027 | 0.849 | 1.2% | 71.3% |
| AZ6 | 26 | 712 | – 5.400 | 0.049 | 0.033 | 0.684 | 1.1% | 80.0% |

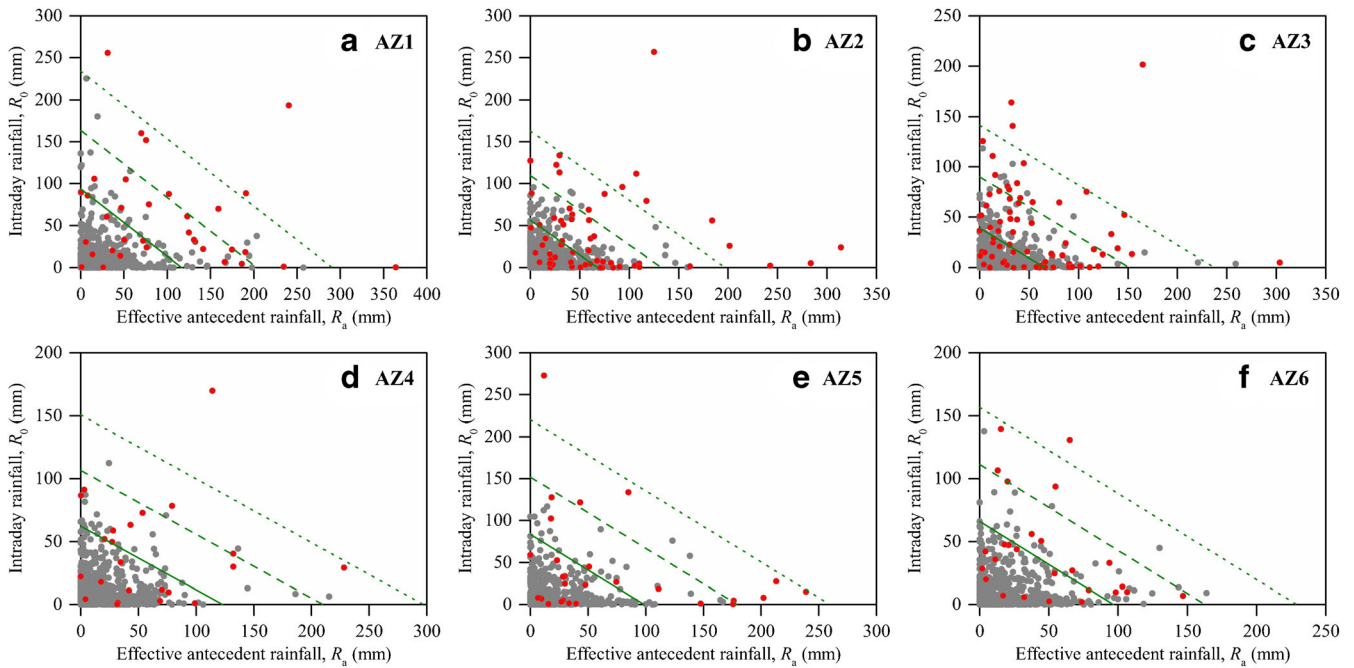


Fig. 7 The intraday rainfall versus effective antecedent rainfall conditions for landslide days (red dots) and non-landslide days (gray dots) and the 10% (solid line), 50% (dashed line), and 90% (dotted line) probability thresholds derived from logistic regression for each alert zone using the calibration dataset

relative significance of antecedent rainfall. In cooler zones or in areas where the permeability of soil is lower, soil water brought by antecedent rainfall can be maintained for a longer period and β would be higher.

Following Eq. (4), the probability of landslide occurrence was computed for each data point in Fig. 7. Theoretically, the probability corresponding to the lower limit of rainfall conditions associated with LDs (p_L) should be equal to 0%, and that

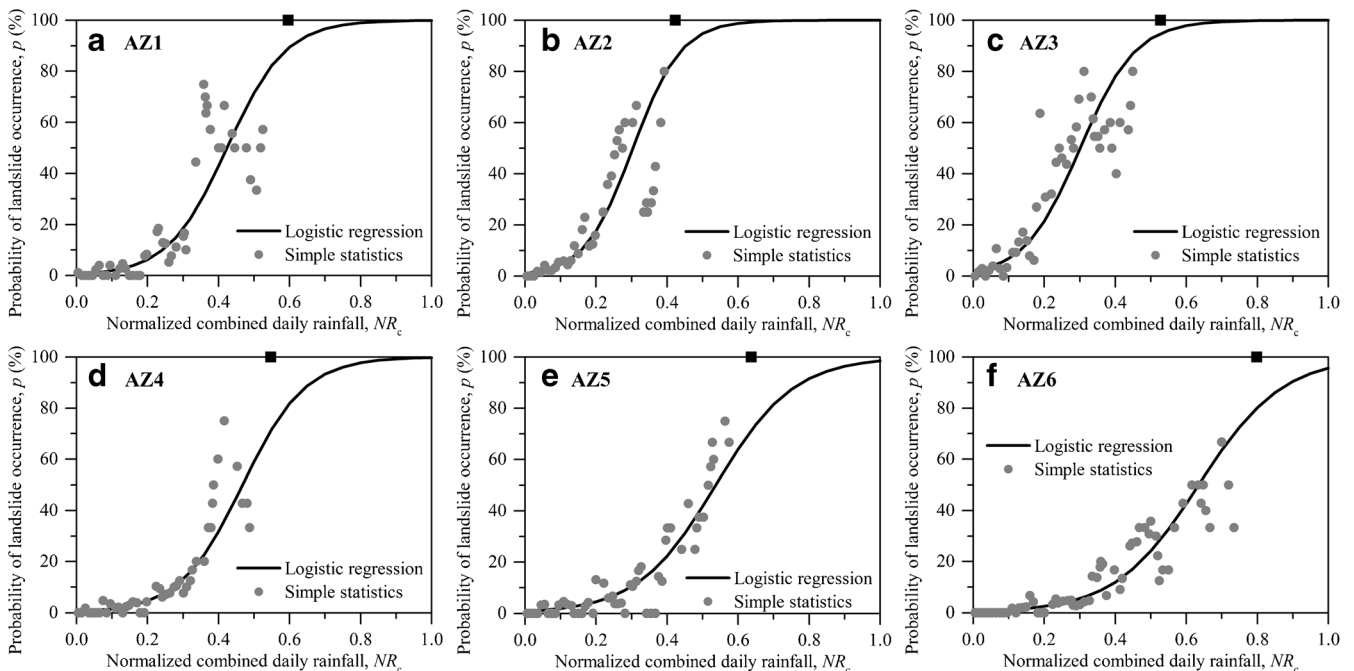


Fig. 8 The relationship between probability of landslide occurrence (p) and normalized combined daily rainfall (NR_c) calculated with the calibration dataset. The gray points represent data obtained by simple statistics, and the black curve was derived from Eq. (7). The maximum NR_c value associated with the non-landslide days and the corresponding theoretical probability (100%) is also shown (the black square)

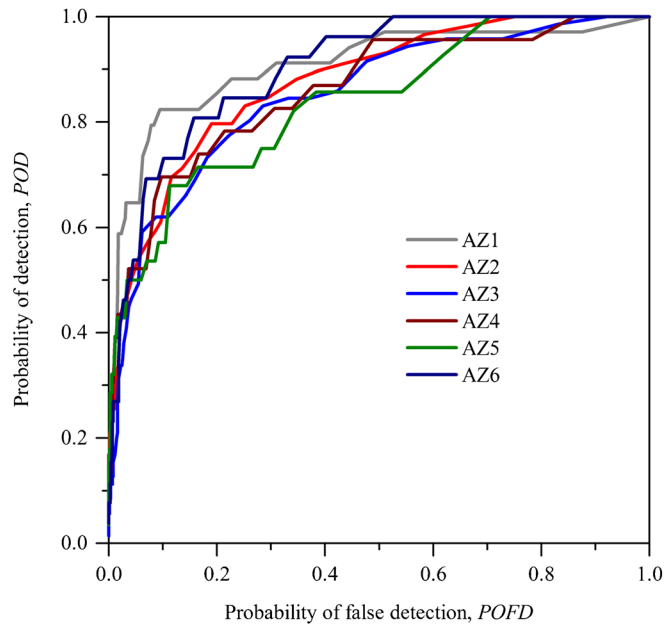


Fig. 9 The ROC curve for each alert zone

corresponding to the upper limit of rainfall conditions associated with NLDs (p_H) should equal 100%. However, the calculated value for p_H is significantly different from the theoretical one in the

study region except AZ3 (Table 3). In terms of p_L , the calculated value approximates the theoretical one. To investigate the reason for the difference, we compared the probability derived from

Table 4 Skill scores corresponding to different probability thresholds for AZ1

| Probability threshold | Classification of the results | | | | Skill scores | | | HK | δ | Λ |
|-----------------------|-------------------------------|------|------|------|--------------|--------------|--------------|--------------|--------------|-----------|
| | TP | FP | TN | FN | POD | $POFD$ | $POFA$ | | | |
| 1% | 33 | 377 | 310 | 1 | <i>0.971</i> | 0.549 | 0.920 | 0.422 | 0.550 | -0.349 |
| 2% | 30 | 184 | 503 | 4 | 0.882 | 0.268 | 0.860 | 0.615 | 0.293 | -0.179 |
| 3% | 28 | 120 | 567 | 6 | 0.824 | 0.175 | 0.811 | 0.649 | 0.248 | -0.137 |
| 4% | 28 | 87 | 600 | 6 | 0.824 | 0.127 | 0.757 | 0.697 | 0.217 | -0.092 |
| 5% | 28 | 69 | 618 | 6 | 0.824 | 0.100 | 0.711 | <i>0.723</i> | <i>0.203</i> | -0.064 |
| 6% | 27 | 59 | 628 | 7 | 0.794 | 0.086 | 0.686 | 0.708 | 0.223 | -0.067 |
| 7% | 27 | 54 | 633 | 7 | 0.794 | 0.079 | 0.667 | 0.716 | 0.220 | -0.057 |
| 8% | 25 | 48 | 639 | 9 | 0.735 | 0.070 | 0.658 | 0.665 | 0.274 | -0.089 |
| 9% | 24 | 41 | 646 | 10 | 0.706 | 0.060 | 0.631 | 0.646 | 0.300 | -0.095 |
| 10% | 22 | 38 | 649 | 12 | 0.647 | 0.055 | 0.633 | 0.592 | 0.357 | -0.133 |
| 20% | 20 | 17 | 670 | 14 | 0.588 | 0.025 | 0.459 | 0.563 | 0.413 | -0.103 |
| 30% | 19 | 11 | 676 | 15 | 0.559 | 0.016 | 0.367 | 0.543 | 0.441 | -0.088 |
| 40% | 13 | 8 | 679 | 21 | 0.382 | 0.012 | 0.381 | 0.371 | 0.618 | -0.209 |
| 50% | 10 | 6 | 681 | 24 | 0.294 | 0.009 | 0.375 | 0.285 | 0.706 | -0.265 |
| 60% | 8 | 5 | 682 | 26 | 0.235 | 0.007 | 0.385 | 0.228 | 0.765 | -0.307 |
| 70% | 7 | 4 | 683 | 27 | 0.206 | 0.006 | 0.364 | 0.200 | 0.794 | -0.319 |
| 80% | 6 | 1 | 686 | 28 | 0.176 | 0.001 | 0.143 | 0.175 | 0.824 | -0.264 |
| 90% | 4 | 0 | 687 | 30 | 0.118 | <i>0.000</i> | <i>0.000</i> | 0.118 | 0.882 | -0.255 |

Best scores are shown in italics

logistic regression with landslide incidence computed by simple statistics. Initially, we defined the combined daily rainfall (R_c) as a linear combination of R_o and R_a :

$$R_c = R_o + \beta R_a \tag{5}$$

To facilitate the analysis, R_c was then normalized by the maximum value calculated for the corresponding AZ ($R_{c,max}$):

$$NR_c = R_c/R_{c,max} \tag{6}$$

in which NR_c is called normalized combined daily rainfall.

For each AZ, landslide incidence for different rainfall conditions was computed using the following procedure: (i) the variable NR_c was computed for each rainy day; (ii) a series of intervals were defined for NR_c , with the lower limit beginning from 0 and increasing by 0.01 step by step, the length being 0.01, 0.02, 0.05, and 0.1, respectively, when the lower limit lies in the ranges 0–0.1, 0.1–0.2, 0.2–0.4, and 0.4–1.0, and the upper limit lower than the maximum NR_c associated with NLDs; (iii) for each interval, landslide incidence was computed as the ratio between the number of LDs and the number of rainy days, and the corresponding value of NR_c was represented by the average of NR_c values for the rainy days. The reason for the usage of changeable interval lengths is that the number of rainy days decreases with the increase in NR_c .

The landslide incidence and NR_c data obtained from the procedure above were finally compared with the logistic regression model:

$$NR_c \cdot R_{c,max} = \frac{1}{\alpha_1} \left(\ln \left(\frac{p}{1-p} \right) - \alpha_0 \right) \tag{7}$$

which is a combination of Eqs. (4)–(6), as shown in Fig. 8. The maximum NR_c value associated with NLDs and the corresponding theoretical probability (100%) is also exhibited in Fig. 8. It suggests that the relationship expressed as Eq. (7) is generally consistent with the statistical data, whereas they agree with each other much better for the smaller NR_c (or p) values than for the larger values. This probably results from two factors. Firstly, there are a very small number of rainy days available for larger NR_c values, leading to a significant fluctuation in the corresponding statistical landslide incidence. Additionally, the number of NLDs is one order of magnitude larger than that of LDs, and thus the rainfall data associated with NLDs exerted a more important influence on the determination of the regression coefficients summarized in Table 3. This might also explain why the computed value of p_H is closest to the theoretical value for the alert zone having the greatest number of LDs (AZ3).

Determination of the optimal rainfall threshold

Despite the uncertainty associated with logistic regression as discussed above, it is still potential to employ this method to define the rainfall threshold, because in this research the threshold was determined by ROC analysis rather than using a specific probability level. The 4 contingencies (TP , FP , TN , and FN) and

Table 5 The optimal rainfall threshold and the corresponding skill scores

| Alert zone | Probability threshold | | Classification of the results | | | | Skill scores | | | δ | \wedge |
|------------|-----------------------|------|-------------------------------|------|--------|--------|--------------|-------|-------|----------|----------|
| | TP | FP | TN | FN | $POFD$ | $POFA$ | HK | POD | | | |
| AZ1 | 27 | 54 | 633 | 7 | 0.079 | 0.667 | 0.716 | 0.794 | 0.220 | –0.057 | |
| AZ2 | 43 | 95 | 591 | 16 | 0.138 | 0.688 | 0.590 | 0.729 | 0.304 | –0.133 | |
| AZ3 | 52 | 113 | 508 | 19 | 0.182 | 0.685 | 0.550 | 0.732 | 0.324 | –0.151 | |
| AZ4 | 15 | 61 | 650 | 8 | 0.086 | 0.803 | 0.566 | 0.652 | 0.358 | –0.196 | |
| AZ5 | 19 | 97 | 612 | 9 | 0.137 | 0.836 | 0.542 | 0.679 | 0.349 | –0.212 | |
| AZ6 | 18 | 52 | 660 | 8 | 0.073 | 0.743 | 0.619 | 0.692 | 0.316 | –0.145 | |

Table 6 The number of the three classes of landslide events with different magnitudes during the 2003–2010 period

| Alert zone | Small | Intermediate | Large |
|------------|-------|--------------|-------|
| AZ1 | 11 | 13 | 10 |
| AZ2 | 31 | 15 | 13 |
| AZ3 | 36 | 23 | 12 |
| AZ4 | 11 | 9 | 3 |
| AZ5 | 13 | 9 | 6 |
| AZ6 | 16 | 9 | 1 |

related skill scores (*POD*, *POFD*, *POFA*, *HK*, and δ) were computed for diverse probability thresholds. The ROC curves are shown in Fig. 9. The *AUC* values change between 0.839 and 0.907 and lie within the 95% confidence intervals, indicating that the prediction capability of the logistic regression model is acceptable.

Taking AZ1 for instance, the skill scores for different probability thresholds are listed in Table 4. The number of true positives (*TP*) is maximized at the 1% probability threshold, and it is the same for the *POD* skill score. However, there are a great number of false alarms (*FP*), leading to high values for the *POFD* and *POFA* skill scores (0.549 and 0.920, respectively). When the threshold increases, both *POFD* and *POFA* decrease for reduced false alarms and get the minimum value (0) at the 90% probability threshold. Nonetheless, *POD* is dramatically low (0.118) in this case owing to the existence of numerous missed alarms (*FN*). In terms of the *HK* skill score, it measures the difference between *POD* and *POFD* and is mathematically $\sqrt{2}$ times the perpendicular distance between the (*POFD*, *POD*) data point and the diagonal line representing random guess (the green line in Fig. 5). Therefore, along the ROC curve, the point associated with the maximum value of *HK* is farthest to the diagonal line. Concerning the convex feature of the ROC curve (Fig. 9), the point having the maximum distance to the diagonal line is usually closest to the perfect classification (the upper left corner of the ROC graph). Consequently, the *HK* skill score and the Euclidean distance δ reach their optimal values at

approximate rainfall thresholds ($p = 5\%$). As a linear combination of *HK*, δ , and *POFA*, the Λ index is maximized at a slightly higher threshold ($p = 7\%$), which was finally selected as the optimal threshold.

Table 5 summarizes the optimal rainfall threshold and the corresponding skill scores for each alert zone. The best performance is obtained at the probability levels $4\% \leq p \leq 10\%$. The corresponding *POD* skill score lies in the range $0.65 \leq POD \leq 0.79$, which is close to the values reported in literature (Staley et al. 2013; Giannecchini et al. 2016). Considering that more than one landslide was recorded in some landslide days, the proportion of landslides that were triggered by rainfall conditions exceeding the optimal threshold is higher than the value of *POD* and varies between 0.71 and 0.95. Although the *POFA* skill score has been concerned to determine the optimal threshold, it still has a relatively high value ($0.67 \leq POFA \leq 0.84$), especially in comparison with *POFD* ($0.07 \leq POFD \leq 0.18$). This most likely results from the larger number of NLDs than that of LDs.

Rainfall thresholds for different warning levels

Table 6 lists the number of the three classes of events with different magnitudes in each AZ during the 2003–2010 period. It suggests that about half of the landslide events are small, which on average occur 1 to 5 times a year, while intermediate events and large events account for 32.4% and 18.7%, respectively. The distribution of the triggering rainfall for each class of events is shown in Fig. 10. Despite the significant variation in most groups of rainfall data, in terms of the median and the third quartile, the triggering rainfall generally increases with landslide magnitude. However, there are two exceptions. One is the median triggering rainfall for intermediate landslide events is lower than that for small events in AZ4. The other is the triggering rainfall for the unique large landslide event is smaller than that for most small and intermediate events. Inspection of the data indicates that these exceptions are probably attributed to two factors: (i) the relatively poor representativeness of the rainfall data in AZ4 and AZ6, and (ii) the incompleteness in landslide records. Undoubtedly, these problems also exist in the data associated with the other AZs. Nevertheless, they are not as

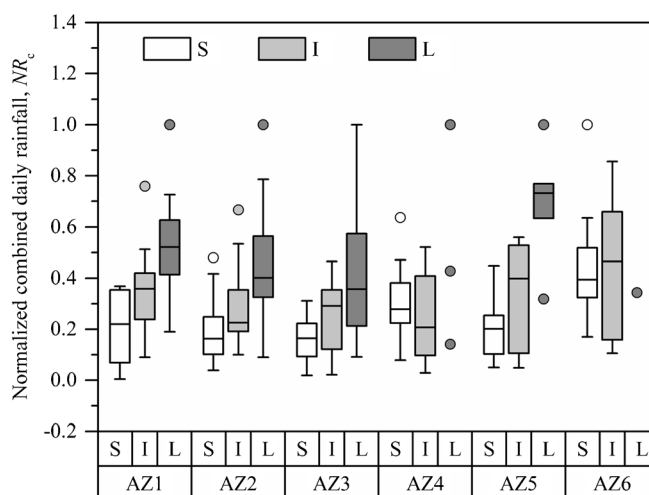


Fig. 10 Comparison of the triggering rainfall for each class of landslide events during the 2003–2010 period (S, small event; I, intermediate event; L, large event)

Table 7 Performance of the warning model with different combinations of probability thresholds for AZ1 during the period 2003–2010

| Probability thresholds | | Classification of the results | | | | | | | Performance indicators | | | | | | |
|------------------------|--------|-------------------------------|------|------|------|------|-------|-------|------------------------|----------|-----------|-----------|--------|--------|----------|
| RI_1 | RI_2 | RT_3 | TN | CP | FA | MA | Gre | Yel | Red | P_{tr} | $FA + MA$ | I_{eff} | HR_L | PR_w | P_{SM} |
| 1% | 30% | 90% | 310 | 393 | 11 | 7 | 326 | 377 | 17 | 1 | 18 | 0.96 | 0.98 | 0.97 | 0.002 |
| 2% | 30% | 90% | 503 | 200 | 11 | 7 | 517 | 186 | 17 | 1 | 18 | 0.92 | 0.97 | 0.95 | 0.005 |
| 3% | 30% | 90% | 567 | 136 | 11 | 7 | 580 | 123 | 17 | 1 | 18 | 0.88 | 0.95 | 0.93 | 0.006 |
| 4% | 30% | 90% | 600 | 103 | 11 | 7 | 613 | 90 | 17 | 1 | 18 | 0.85 | 0.94 | 0.90 | 0.008 |
| 5% | 30% | 90% | 618 | 85 | 11 | 7 | 631 | 72 | 17 | 1 | 18 | 0.83 | 0.92 | 0.89 | 0.010 |
| 6% | 30% | 90% | 628 | 75 | 11 | 7 | 641 | 62 | 17 | 1 | 18 | 0.81 | 0.91 | 0.87 | 0.011 |
| 7% | 30% | 90% | 633 | 70 | 11 | 7 | 646 | 57 | 17 | 1 | 18 | 0.80 | 0.91 | 0.86 | 0.011 |
| 8% | 30% | 90% | 639 | 64 | 11 | 7 | 650 | 53 | 17 | 1 | 18 | 0.78 | 0.90 | 0.85 | 0.012 |
| 9% | 30% | 90% | 646 | 57 | 11 | 7 | 657 | 46 | 17 | 1 | 18 | 0.76 | 0.89 | 0.84 | 0.013 |
| 10% | 30% | 90% | 649 | 54 | 11 | 7 | 659 | 44 | 17 | 1 | 18 | 0.75 | 0.89 | 0.83 | 0.014 |
| 7% | 10% | 90% | 633 | 45 | 38 | 5 | 647 | 31 | 42 | 1 | 43 | 0.51 | 0.90 | 0.54 | 0.011 |
| 7% | 20% | 90% | 633 | 64 | 17 | 7 | 645 | 52 | 23 | 1 | 24 | 0.73 | 0.90 | 0.79 | 0.011 |
| 7% | 30% | 90% | 633 | 70 | 11 | 7 | 646 | 57 | 17 | 1 | 18 | 0.80 | 0.91 | 0.86 | 0.011 |
| 7% | 40% | 90% | 633 | 70 | 8 | 10 | 646 | 57 | 17 | 1 | 18 | 0.80 | 0.88 | 0.90 | 0.011 |
| 7% | 50% | 90% | 633 | 69 | 6 | 13 | 645 | 57 | 16 | 3 | 19 | 0.78 | 0.84 | 0.92 | 0.034 |
| 7% | 60% | 90% | 633 | 68 | 5 | 15 | 644 | 57 | 16 | 4 | 20 | 0.77 | 0.82 | 0.93 | 0.045 |
| 7% | 70% | 90% | 633 | 68 | 4 | 16 | 644 | 57 | 15 | 5 | 20 | 0.77 | 0.81 | 0.94 | 0.057 |
| 7% | 80% | 90% | 633 | 70 | 1 | 17 | 643 | 60 | 13 | 5 | 18 | 0.80 | 0.80 | 0.99 | 0.057 |
| 7% | 30% | 40% | 633 | 70 | 11 | 7 | 649 | 54 | 9 | 9 | 18 | 0.80 | 0.91 | 0.86 | 0.102 |
| 7% | 30% | 50% | 633 | 70 | 11 | 7 | 648 | 55 | 11 | 7 | 18 | 0.80 | 0.91 | 0.86 | 0.080 |
| 7% | 30% | 60% | 633 | 70 | 11 | 7 | 648 | 55 | 12 | 6 | 18 | 0.80 | 0.91 | 0.86 | 0.068 |
| 7% | 30% | 70% | 633 | 70 | 11 | 7 | 647 | 56 | 13 | 5 | 18 | 0.80 | 0.91 | 0.86 | 0.057 |
| 7% | 30% | 80% | 633 | 70 | 11 | 7 | 648 | 55 | 16 | 2 | 18 | 0.80 | 0.91 | 0.86 | 0.023 |
| 7% | 30% | 90% | 633 | 70 | 11 | 7 | 646 | 57 | 17 | 1 | 18 | 0.80 | 0.91 | 0.86 | 0.011 |

Best scores are shown in *italics*

severe as the case in AZ4 and AZ6 and have not influenced the positive relationship between landslide magnitude and triggering rainfall.

To define the three thresholds (RT_1 , RT_2 , and RT_3) for the four warning levels, sensitivity of the four performance indicators (I_{eff} , HR_L , PP_W , P_{SM}) to each threshold was analyzed. Taking AZ1 for example, the value of the indicators was computed for different probability combinations. The results are shown in Table 7. The number of correct predictions (CP) increases with the increase in RT_1 , while the number of true negatives (TN) exhibits an opposite trend. The change in RT_1 has no influence on the number of false alarms (FA) and missed alarms (MA). Correspondingly, the maximum value for I_{eff} , HR_L , and PP_W is obtained at the smallest RT_1 ($p = 1\%$). In terms of the P_{ur} indicator, it is also optimized at the 1% probability threshold for RT_1 . It seems that the smaller the probability for RT_1 , the better the model performance. However, further analysis reveals that the variation in all the four indicators is simply induced by the variation in the number of CP while RT_2 and RT_3 remain unchanged. Comparison between the two classes of criteria illustrated in Fig. 6 manifests that landslide events pertaining to CP come from two sources: events with no prediction error but with landslide activity ($Gre - TN$) and events with minor error (Yel). In other words, a mathematical relationship exists among the number of CP , Yel , Gre , and TN : $CP = Yel + (Gre - TN)$. When the probability for RT_1 decreases from 10 to 1%, the number of ($Gre - TN$) rises from 10 to 16, while the number of Yel grows from 44 to 377, suggesting that the increase in the number of CP is primarily induced by the increase in the number of the events with minor prediction error. That is to say, if the 1% probability threshold is used for RT_1 in the operational early warning system, a great number of WL_1 would be issued for the days without landslide occurrence, and the reliability of the system would be affected. Concerning this issue, the optimal threshold determined by ROC analysis was finally chosen as RT_1 , i.e., $p = 7\%$. Compared with the 1% probability threshold, the proportion of the events classified as Yel in those pertaining to CP decreases from 0.96 to 0.81.

In the aspect of the RT_2 threshold, when it becomes larger, fewer events are classified as EA while a greater number of MA are generated. Since the number of TN keeps constant, the number of CP is just related to the total number of incorrect predictions ($EA + MA$). The number of Pur increases marginally with the increase in RT_2 . Correspondingly, the I_{eff} , HR_L , PP_W , and P_{SM} indicators change in different ways when RT_2 is altered and arrive at their optimal values at different probability thresholds: $p = 30\%$, 40% , and 80% for I_{eff} , $p = 30\%$ for HR_L , $p = 80\%$ for PP_W , and $p \leq 40\%$ for P_{SM} . In regard to RT_3 , the number of CP , EA , and MA is insensitive to this threshold, while the number of Pur decreases drastically when RT_3 increases. Therefore, a high threshold ($p = 90\%$) is required to minimize P_{SM} . The determination of RT_2 still remains a problem. Inspection of Table 7 indicates that the number of Pur remains at the minimum when RT_2 varies in a relatively wide range ($p = 10 - 40\%$), in which the optimal value of $EA + MA$ is 18 and it remains the smallest when more thresholds are considered. As a consequence, the optimal RT_2 threshold can be identified as the one which primarily minimizes the occurrences with severe error (Pur) and then generates the smallest possible incorrect predictions ($EA + MA$). If more than one threshold is selected following this criterion, the lowest one, namely the most conservative one, will be finally utilized. For instance, although the 30% and 40% probability thresholds provide equivalent values for Pur and for $EA + MA$ (Table 7), the lower threshold generates less missed alarms and more false alarms and was ultimately selected for RT_2 .

Table 8 The RT_1 , RT_2 , and RT_3 thresholds and the corresponding performance

| Alert zone | Probability thresholds | | | Classification of the results | | | | | | | Performance indicators | | | | |
|------------|------------------------|--------|--------|-------------------------------|------|------|------|-------|-------|-------|------------------------|-----------|--------|--------|----------|
| | RT_1 | RT_2 | RT_3 | TN | CP | EA | MA | Gre | Yel | Red | Pur | I_{eff} | HR_L | PP_W | P_{SM} |
| AZ1 | 7% | 30% | 90% | 633 | 70 | 11 | 7 | 646 | 57 | 17 | 1 | 0.80 | 0.91 | 0.86 | 0.011 |
| AZ2 | 9% | 45% | 95% | 591 | 133 | 8 | 13 | 614 | 110 | 19 | 2 | 0.86 | 0.91 | 0.94 | 0.013 |
| AZ3 | 10% | 45% | 95% | 508 | 159 | 11 | 14 | 547 | 120 | 21 | 4 | 0.86 | 0.92 | 0.94 | 0.022 |
| AZ4 | 5% | 30% | 90% | 650 | 72 | 5 | 7 | 661 | 61 | 11 | 1 | 0.86 | 0.91 | 0.94 | 0.012 |
| AZ5 | 4% | 45% | 75% | 612 | 116 | 2 | 7 | 626 | 102 | 8 | 1 | 0.93 | 0.94 | 0.98 | 0.008 |
| AZ6 | 7% | 55% | 85% | 660 | 67 | 4 | 7 | 673 | 54 | 10 | 1 | 0.86 | 0.91 | 0.94 | 0.013 |

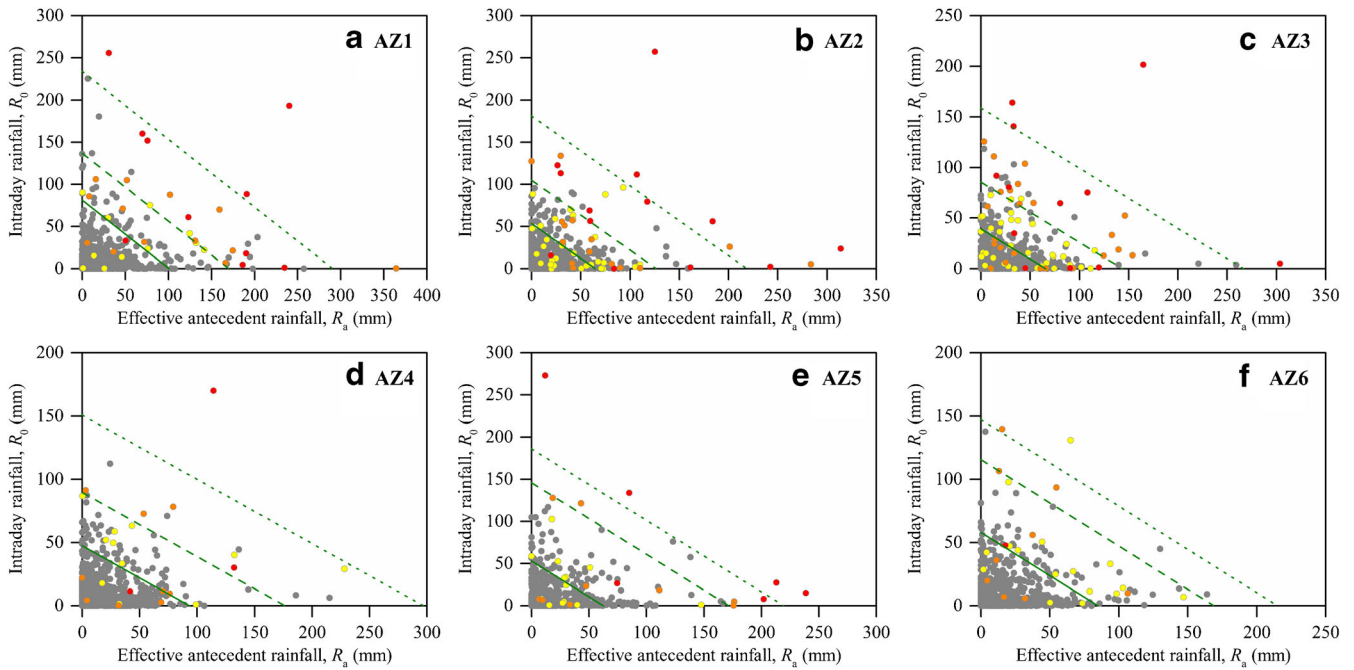


Fig. 11 The intraday rainfall versus effective antecedent rainfall conditions for the four classes of landslide events during the 2003–2010 period and the three thresholds for the warning model. The gray, yellow, orange, and red dots correspond to no landslides, small events, intermediate events, and large events, respectively. The solid, dashed, and dotted lines represent the RT_1 , RT_2 , and RT_3 thresholds, respectively

Table 9 Performance of the rainfall thresholds for landslide occurrence for the period 2011–2015

| Alert zone | Number of days Rainy day | LD | Classification of the results | | | | Skill scores | | | |
|------------|-----------------------------|----|-------------------------------|----|-----|----|--------------|-------|-------|-------|
| | | | TP | FP | TN | FN | POD | POFD | POFA | HK |
| AZ1 | 476 | 17 | 13 | 36 | 423 | 4 | 0.765 | 0.078 | 0.735 | 0.686 |
| AZ2 | 475 | 14 | 14 | 63 | 398 | 0 | 1.000 | 0.137 | 0.818 | 0.863 |
| AZ3 | 480 | 49 | 35 | 70 | 361 | 14 | 0.714 | 0.162 | 0.667 | 0.552 |
| AZ4 | 479 | 2 | 2 | 68 | 409 | 0 | 1.000 | 0.143 | 0.971 | 0.857 |
| AZ5 | 500 | 15 | 11 | 66 | 419 | 4 | 0.733 | 0.136 | 0.857 | 0.597 |
| AZ6 | 513 | 2 | 2 | 49 | 462 | 0 | 1.000 | 0.096 | 0.961 | 0.904 |

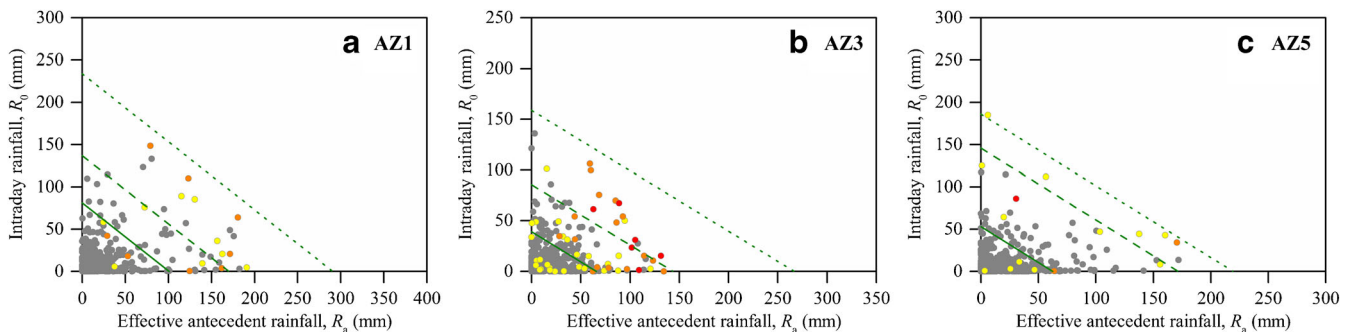


Fig. 12 The intraday rainfall versus effective antecedent rainfall conditions for the four classes of landslide events during the period 2011–2015 and the three thresholds for the warning model. The gray, yellow, orange, and red dots correspond to no landslides, small events, intermediate events, and large events, respectively. The solid, dashed, and dotted lines represent the RT_1 , RT_2 , and RT_3 thresholds, respectively

Table 10 Performance of the rainfall thresholds for different warning levels for the period 2011–2015

| Alert zone | Number of events | | Classification of the results | | | | | | Performance indicators | | | | | |
|------------|------------------|----|-------------------------------|-----|----|----|-----|-----|------------------------|-----|-----------|--------|--------|----------|
| | S | L | TN | CP | FA | MA | Gre | Yel | Red | Pur | I_{eff} | HR_L | PP_w | P_{SM} |
| AZ1 | 9 | 0 | 423 | 39 | 10 | 4 | 429 | 33 | 14 | 0 | 0.74 | 0.91 | 0.80 | 0.000 |
| AZ3 | 27 | 16 | 361 | 101 | 6 | 12 | 379 | 83 | 16 | 2 | 0.85 | 0.89 | 0.94 | 0.017 |
| AZ5 | 12 | 2 | 419 | 77 | 2 | 2 | 424 | 72 | 3 | 1 | 0.95 | 0.97 | 0.97 | 0.012 |

To obtain the optimal RT_2 and RT_3 thresholds for each alert zone, more probability combinations were employed. Using the probability for RT_2 and RT_3 varying in the range $20\% \leq p \leq 90\%$ and $35\% \leq p \leq 95\%$, respectively, with an interval of 5%, a total of 117 combinations were considered. The optimal combination was identified following the criteria discussed above and is listed in Table 8. For AZ1 and AZ5, the values of Pur and $FA + MA$ are minimized at the selected probability combination, while for the other AZs the value of $FA + MA$ is the second smallest among the 117 combinations. Inspection of Table 8 shows that the optimal probability for RT_1 , RT_2 , and RT_3 lies within the ranges $4\% \leq p \leq 10\%$, $30\% \leq p \leq 55\%$, and $75\% \leq p \leq 95\%$, respectively, which coincidentally is approximate to the probability thresholds used for four warning levels in literature ($p = 10\%$, 50% , and 90% , respectively)(Zhuang et al. 2014; Huang et al. 2015). The thresholds are also illustrated in the R_o-R_a space (Fig. 11).

Validation of the rainfall thresholds

To validate the thresholds, landslide and rainfall data during the monsoon season of the period 2011–2015 were analyzed. Table 9 lists the number of rainy days and LDs in each zone. Using the thresholds summarized in Table 5, the four contingencies and related skill scores were computed, as shown in Table 9. Comparison with skill scores derived from the calibration dataset, values of POD are generally higher for the validation dataset while values of $POFD$ are just slightly different. Consequently, the HK skill score becomes higher. Nevertheless, $POFA$ is also higher for most zones, especially for AZ4 and AZ6, where only 2 days were recorded with landslide occurrence during the 5-year period. Investigation of the rainfall data reveals that the average daily precipitation for rainy days in the monsoon season of the period 2011–2015 (11.9 mm) is lower than that for the period 2003–2010 (12.6 mm). Therefore, the proportion of LDs has decreased. However, the decrease is much more drastic in AZ2, AZ4, and AZ6 than in the other zones, which means that the landslide records are probably severely incomplete in these zones because the local governments have put more effort into geohazard mitigation and a large portion of landslides were not recorded considering that no damage was caused, leading to higher $POFA$ values. As a result, using skill scores from the remaining zones to evaluate performance of the thresholds is more reliable, which shows that the skill scores derived from the validation dataset are comparable to that derived from the calibration dataset, with POD , $POFD$, and $POFA$ changing in the ranges 0.71–0.77, 0.08–0.16, and 0.67–0.86, respectively.

The validation dataset from AZ1, AZ3, and AZ5 were further used to evaluate the RT_1 , RT_2 , and RT_3 thresholds, as illustrated in Fig. 12. Table 10 summarizes the classification of the results and corresponding performance indicators. Comparison between Table 8 and Table 10 shows that although the I_{eff} and PP_w indicators slightly decrease for AZ1 in the period 2011–2015, there is not much difference between the prediction performance for the same zone in the two periods, indicating the robustness of the thresholds and the potential of them for practical use.

Conclusions

In this study, daily rainfall data from 6 gauges and landslide data including 1836 records in Dazhou during the period of 2003–2015 were systematically analyzed. The study region was divided into six alert zones. For each zone, the rainfall threshold for landslide

activity was defined using logistic regression, ROC analysis, and a procedure for performance evaluation of models with multiple warning levels with dataset for the period 2003–2010, and was further validated with dataset for the period 2011–2015. The following conclusions can be drawn:

1. The probability of landslide occurrence computed by logistic regression was compared with the value counted by simple statistics for various rainfall conditions. The difference between the two values might be relatively large when the probability is high. Therefore, it is highly suggested that the rainfall threshold be defined using a rigorous methodology such as ROC analysis rather than using a predetermined probability level.
2. Diverse threshold combinations were used to evaluate performance of the model with four warning levels, and the best-performing combination was determined. Performance of the threshold for the validation dataset is comparable to that for the calibration dataset, and the efficiency index varies in the range $0.74 \leq I_{\text{eff}} \leq 0.95$. Considering the large geographical area of each alert zone, rainfall data measured by the unique gauge is insufficient to represent rainfall conditions in the whole zone. Nonetheless, performance of the warning model is acceptable. This is probably because rainfall parameters used in this study are associated with a longer period (8 days) of rainfall data, in which case the representativeness of the gauge might be better than the case that event rainfall data are used. Hence, the procedure for threshold determination employed in this research is expected to be applicable to other poorly gauged areas.>

Many factors, such as geology, topography, and vegetation, also play important roles in slope stability. It is impossible to predict possible landslide activities merely using rainfall data. Consequently, it should be emphasized that the rainfall threshold should be used in combination with landslide susceptibility zonation techniques, in which the other factors are considered, in operational regional warning systems.

Acknowledgments

The authors sincerely appreciate the data supply of the Sichuan Provincial Environmental Monitoring Center and the valuable comments from the anonymous reviewers.

Funding information

This work was financially supported by the National Key R&D Program of China (grant No. 2018YFC1505503), the Science and Technology Support Project of Sichuan Province (grant No. 2015SZ0214), and the 135 Strategic Program of Institute of Mountain Hazards and Environment, CAS (grant No. SDS-135-1705).

References

Aleotti P (2004) A warning system for rainfall-induced shallow failures. *Eng Geol* 73:247–265
 Baum RL, Godt JW (2010) Early warning of rainfall-induced shallow landslides and debris flows in the USA. *Landslides* 7:259–272
 Bezak N, Šraj M, Mikoš M (2016) Copula-based IDF curves and empirical rainfall thresholds for flash floods and rainfall-induced landslides. *J Hydrol* 541:272–284
 Bruce JP, Clark RH (1966) *Introduction to hydrometeorology*. Pergamon, Oxford

Bui DT, Pradhan B, Lofman O, Revhaug I, Dick ØB (2013) Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. *Nat Hazards* 66:707–730
 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
 Chang KT, Chiang SH, Lei F (2008) Analysing the relationship between typhoon-triggered landslides and critical rainfall conditions. *Earth Surf Proc Land* 33:1261–1271
 Chung MC, Tan CH, Chen CH (2017) Local rainfall thresholds for forecasting landslide occurrence: Taipingshan landslide triggered by typhoon Saola. *Landslides* 14:19–33
 Crozier MJ, Eyles RJ (1980) Assessing the probability of rapid mass movement. In: The New Zealand Institution of Engineers—Proceedings of Technical Groups (ed) Proceedings of the 3rd Australia–New Zealand Conference on Geomechanics, Wellington, Vol 2, pp 47–51
 Elias GU (2016) Establishing rainfall frequency contour lines as thresholds for rainfall-induced landslides in Tegucigalpa, Honduras, 1980–2005. *Nat Hazards* 82:2107–2132
 Frattini P, Crosta G, Sosio R (2009) Approaches for defining thresholds and return periods for rainfall-triggered shallow landslides. *Hydrol Process* 23(10):1444–1460
 Gao L, Zhang LM, Cheung RWM (2018) Relationships between natural terrain landslide magnitudes and triggering rainfall based on a large landslide inventory in Hong Kong. *Landslides* 15:727–740
 Gariano SL, Brunetti MT, Iovine G, Melillo M, Peruccacci S, Terranova O, Vennari C, Guzzetti F (2015) Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy. *Geomorphology* 228:653–665
 Giannacchini R, Galanti Y, Avanzi GD, Barsanti M (2016) Probabilistic rainfall thresholds for triggering debris flows in a human-modified landscape. *Geomorphology* 257:94–107
 Glade T, Nadim F (2014) Early warning systems for natural hazards and risks. *Nat Hazards* 70:1669–1671
 Glade T, Crozier M, Smith P (2000) Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical “Antecedent Daily Rainfall Model”. *Pure Appl Geophys* 157:1059–1079
 Guzzetti F, Peruccacci S, Rossi M, Stark CP (2007) Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorol Atmos Phys* 98:239–267
 Guzzetti F, Peruccacci S, Rossi M, Stark CP (2008) The rainfall intensity–duration control of shallow landslides and debris flows: an update. *Landslides* 5:3–17
 Hanssen AW, Kuipers WJA (1965) On the relationship between the frequency of rain and various meteorological parameters. *Koninklijk Nederlands Meteorologisch Instituut Meded Verh* 81:2–15
 Huang J, Ju NP, Liao YJ, Liu DD (2015) Determination of rainfall thresholds for shallow landslides by a probabilistic and empirical method. *Nat Hazards Earth Syst Sci* 15:2715–2723
 Kanjanakul C, Chub-uppakarn T, Chalermyanont T (2016) Rainfall thresholds for landslide early warning system in Nakhon Si Thammarat. *Arab J Geosci* 9:584
 Kanungo DP, Sharma S (2014) Rainfall thresholds for prediction of shallow landslides around Chamoli–Joshimath region, Garhwal Himalayas, India. *Landslides* 11:629–638
 Keefer DK, Wilson RC, Mark RK, Brabb EE, Brown WM, Ellen SD, Harp EL, Wieczorek GF, Alger CS, Zarkin RS (1987) Real-time landslide warning during heavy rainfall. *Science* 238:921–925
 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
 Lainas S, Sabatakakis N, Koukis G (2016) Rainfall thresholds for possible landslide initiation in wildfire-affected areas of western Greece. *Bull Eng Geol Environ* 75:883–896
 Liu D, Zhu C, Tian Y (2014) Spatial-temporal distribution of geohazards in Dazhou and cause analysis. *Acta Geol Sichuan* 34(Suppl):31–35 (In Simplified Chinese)
 Ma T, Li C, Lu Z, Wang B (2014) An effective antecedent precipitation model derived from the power-law relationship between landslide occurrence and rainfall level. *Geomorphology* 216:187–192
 Ma T, Li C, Lu Z, Bao Q (2015) Rainfall intensity–duration thresholds for the initiation of landslides in Zhejiang Province, China. *Geomorphology* 245:193–206
 Mathew J, Babu DG, Kundu S, Kumar KV, Pant CC (2014) Integrating intensity–duration-based rainfall threshold and antecedent rainfall-based probability estimate towards generating early warning for rainfall-induced landslides in parts of the Garhwal Himalaya, India. *Landslides* 11:575–588
 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* 13:165–172
 Peres DJ, Cancelliere A (2014) Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach. *Hydrol Earth Syst Sci* 18:4913–4931
 Piciullo L, Gariano SL, Melillo M, Brunetti MT, Peruccacci S, Guzzetti F, Calvello M (2017) Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. *Landslides* 14:995–1008

- Posner AJ, Georgakakos KP (2015) Soil moisture and precipitation thresholds for real-time landslide prediction in El Salvador. *Landslides* 12:1179–1196
- Rosi A, Peternel T, Jemec-Auflič M, Komac M, Segoni S, Casagli N (2016) Rainfall thresholds for rainfall-induced landslides in Slovenia. *Landslides* 13:1571–1577
- Rossi M, Luciani S, Valigi D, Kirschbaum D, Brunetti MT, Peruccacci S, Guzzetti F (2017) Statistical approaches for the definition of landslide rainfall thresholds and their uncertainty using rain gauge and satellite data. *Geomorphology* 285:16–27
- Segoni S, Rosi A, Rossi G, Catani F, Casagli 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
- Segoni S, Lagomarsino D, Fanti R, Moretti S, Casagli N (2015) Integration of rainfall thresholds and susceptibility maps in the Emilia Romagna (Italy) regional-scale landslide warning system. *Landslides* 12:773–785
- Sepúlveda SA, Padilla C (2008) Rain-induced debris and mudflow triggering factors assessment in the Santiago cordilleran foothills, Central Chile. *Nat Hazards* 47:201–215
- Staley DM, Kean JW, Cannon SH, Schmidt KM, Laber JL (2013) Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California. *Landslides* 10:547–562
- Tang XD, Ma WF, Yang QG, Zhang SZ (2010) The remote monitoring system of Dazhou geological disaster based on MAS. *Comput. Tech and Its Appl* (6):148–154 (In Simplified Chinese)
- Vallet A, Varron D, Bertrand C, Fabbri O (2016) A multi-dimensional statistical rainfall threshold for deep landslides based on groundwater recharge and support vector machines. *Nat Hazards* 84:821–849
- 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
- Wu Y, Lan H, Gao X, Li L, Yang Z (2015) A simplified physically based coupled rainfall threshold model for triggering landslides. *Eng Geol* 195:63–69
- Xiang B (2015) Types, genesis and control of geo-hazards in Dachuan, Dazhou. *J Geol Hazards Environ Pres* 26(3):65–72 (In Simplified Chinese)
- Yu B, Wu Y, Chu S (2014) Preliminary study of the effect of earthquakes on the rainfall threshold of debris flows. *Eng Geol* 182:130–135
- Zhuang J, Iqbal J, Peng J, Liu T (2014) Probability prediction model for landslide occurrences in Xi'an, Shaanxi Province, China. *J Mt Sci* 11(2):345–359

H. Yang

Key Laboratory of Mountain Hazards and Earth Surface Process, Chinese Academy of Sciences, Chengdu, 610041, China

H. Yang · P. Su · S. Zhang

Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, 610041, China

F. Wei (✉)

Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, 400714, China
Email: fqwei@imde.ac.cn

Z. Ma · H. Guo

Sichuan Climate Center, Chengdu, 610071, China