Landslides (2013) 10:547–562
DOI 10.1007/s10346-012-0341-9
Received: 29 September 2011
Accepted: 7 June 2012
Published online: 27 June 2012
© Springer-Verlag (outside the USA) 2012

Dennis M. Staley · Jason W. Kean · Susan H. Cannon · Kevin M. Schmidt · Jayme L. Laber

Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California

Abstract Rainfall intensity–duration (ID) thresholds are commonly used to predict the temporal occurrence of debris flows and shallow landslides. Typically, thresholds are subjectively defined as the upper limit of peak rainstorm intensities that do not produce debris flows and landslides, or as the lower limit of peak rainstorm intensities that initiate debris flows and landslides. In addition, peak rainstorm intensities are often used to define thresholds, as data regarding the precise timing of debris flows and associated rainfall intensities are usually not available, and rainfall characteristics are often estimated from distant gauging locations. Here, we attempt to improve the performance of existing threshold-based predictions of post-fire debris-flow occurrence by utilizing data on the precise timing of debris flows relative to rainfall intensity, and develop an objective method to define the threshold intensities. We objectively defined the thresholds by maximizing the number of correct predictions of debris flow occurrence while minimizing the rate of both Type I (false positive) and Type II (false negative) errors. We identified that (1) there were statistically significant differences between peak storm and triggering intensities, (2) the objectively defined threshold model presents a better balance between predictive success, false alarms and failed alarms than previous subjectively defined thresholds, (3) thresholds based on measurements of rainfall intensity over shorter duration (≤60min) are better predictors of post-fire debris-flow initiation than longer duration thresholds, and (4) the objectively defined thresholds were exceeded prior to the recorded time of debris flow at frequencies similar to or better than subjective thresholds. Our findings highlight the need to better constrain the timing and processes of initiation of landslides and debris flows for future threshold studies. In addition, the methods used to define rainfall thresholds in this study represent a computationally simple means of deriving critical values for other studies of nonlinear phenomena characterized by thresholds.

Keywords Debris flow · Wildfire · Rainfall · Thresholds · Warning system

Introduction

Rainfall-initiated landslides and debris flows are a common hazard in mountainous terrain throughout the world. As human population grows and landslide-susceptible areas are further developed, the frequency and severity of landslide and debris-flow impacts to populations and infrastructure are likely to increase. A common goal of landslide and debris-flow research is to improve predictions of the occurrence of these events in both space and time. Susceptibility maps and models are typically used to estimate where these events are likely to occur (Carrara et al. 2008; Godt et al. 2008; Cannon et al. 2009; Baum et al. 2010), while critical thresholds are used to define the meteorological, geotechnical or hydrological conditions that, when exceeded, are likely to produce landslides or debris flows (Caine 1980; Wieczorek and Guzzetti 2000; Bracchini and Zannoni 2003; Godt et al. 2006; Guzzetti et al. 2007; Cannon et al. 2008; Guzzetti et al. 2008; Baum and Godt 2010; Cannon et al. 2011).

In general, critical thresholds represent the point at which slight changes in an internal or external variable produces abrupt changes or failures within the affected system (Schumm 1973). For shallow landslides and debris flows, a measurement of rainfall is typically used as the external variable, such as average intensity (Caine 1980; Guzzetti et al. 2008), antecedent rainfall (Martelloni et al. 2012), rainfall intensity as measured over a given duration (Cannon et al. 2008, 2011), or some combination of measures (Aleotti 2004; Godt et al. 2006; Ponziani et al., 2012). The critical value of the threshold is governed at the site- or region-specific scale by physiographic properties such as antecedent conditions, topography, lithology, soil properties and land cover.

Studies aimed at defining the rainfall intensity–duration (ID) thresholds are the most common type of threshold used in landslide and debris-flow studies. They may be defined empirically using historical data (Caine 1980; Aleotti 2004; Godt et al. 2006; Guzzetti et al. 2007; Cannon et al. 2008; Coe et al. 2008; Guzzetti et al. 2008; Cannon et al. 2011), theoretically using physically based models (Montgomery and Dietrich 1994; Iiritano et al. 1998; Godt and McKenna 2008; Baum et al. 2010; Salciarini et al. 2011), or through some combination of empirical and theoretical methods (Crozier 1999). A challenge for applying physically based models is that they require detailed meteorological, hydrological and geotechnical data for accurate model calibration and prediction. These data can be difficult to collect over large geographic areas. Consequently, regional and global predictions of the temporal occurrence of landslides and debris flows are usually made using empirical thresholds derived from more easily obtained data on rainfall and the presence/absence of landslides or debris flows during a given storm (Caine 1980; Larsen and Simon 1993; Godt et al. 2006; Hong et al. 2006; Guzzetti et al. 2007; Cannon et al. 2008; Coe et al. 2008; Guzzetti et al. 2008; Cannon et al. 2011).

Empirical ID thresholds are typically represented by a power law equation in the form of $I=\alpha D^{\beta}$, where I is the rainfall intensity (in mm/h), D is the duration (in hours), and α and β are empirically derived parameters (Caine 1980). The threshold approach to predicting landslide and debris-flow occurrence using a power law equation is based upon two assumptions. First, there is a nonlinear increase in the probability of landslide or debris flow initiation with increasing rainfall intensities. This is represented by the actual threshold value for slide or flow initiation. Below the threshold value, there is a lower probability of debris flow or landslide initiation. At and above the threshold, there is a rapid, nonlinear increase in the likelihood of initiation. In the power law equation, the coefficient α and the exponent β define the location of the critical intensity value. Second, as the duration of rainfall increases, there is a decrease in the intensity of rainfall needed to trigger landslide or debris flow

initiation. In the power law equation, the exponent β defines the rate at which critical intensity decreases with increasing rainfall duration.

Ideally, ID thresholds for the prediction of landslide or debris-flow initiation should meet the following criteria. First, the threshold rainfall intensities should be representative of the rainfall intensities that contribute to the initiation of landslides or debris flows. Second, events should occur at the time critical rainfall ID thresholds have been exceeded or shortly thereafter. Third, the number of instances where events occur below the threshold value (failed alarms) should be minimized. Finally, the number of threshold-exceeding storms that did not produce landslides and debris flows (false alarms) should also be minimized. Two aspects of standard empirical threshold delineation could be improved to achieve thresholds closer to this ideal threshold. First, the data used to derive rainfall threshold conditions for landslides and debris-flow initiation usually do not take into account the actual rainfall conditions that immediately precede the occurrences of these events (hereafter referred to as the triggering conditions) and rely instead upon measurements of peak rainfall intensity during the storm for threshold derivation. Although some studies have evaluated the timing of landslides or debris flows relative to rainfall conditions (Cannon and Ellen 1988; Berti et al. 1999; Chien-Yuan et al. 2005; Gregoretti and Fontana 2007, 2008; McCoy et al. 2010), landslides and debris flows frequently occur in areas with no monitoring equipment and where eyewitness accounts are unavailable or temporally imprecise. A recent review of a global database of rainfall thresholds revealed that only 5.1 % (133) of the 2,626 published landslide and debris-flow events contained timing data accurate to ±12 h (Guzzetti et al. 2008). Given a lack of precise timing data, researchers have often relied upon measures of peak or average rainstorm intensities as representative of the hydrometeorological conditions that initiate landslides and debris flows.

The second improvement would be the manner in which empirically derived ID thresholds are defined. Most studies have traditionally defined thresholds using a subjective approach, and only recently have there been attempts to objectively define rainfall ID thresholds using numerical and statistical methods (Guzzetti et al. 2007, 2008; Brunetti et al. 2010; Montesarchio et al. 2011).A major limitation of the subjective method is that it precludes repeatability of the method and makes it difficult to compare thresholds defined by different researchers (Guzzetti et al. 2007, 2008). The method employed in the subjective definition of ID thresholds typically follows one of the two following strategies. The first strategy defines the lower limit of rainfall intensities above which debris flows or landslides have been found to initiate (e.g., Caine 1980; Godt et al. 2006; Cannon et al. 2008). Hereafter, we refer to this approach as the "lower limit" (LL) method of threshold definition. Thresholds derived for the lower limit represent the most conservative rainfall conditions (e.g., lower rainfall intensities) for the initiation of debris flows, where rainfall intensities that occur above the threshold often do not result in the initiation of landslides or debris flows. The second strategy defines the threshold at the upper limit of rainfall intensities of storms that did not produce a debris flow or landslide (e.g., Larsen and Simon 1993; Cannon et al. 2008). Hereafter, we refer to this approach as the "upper limit" (UL) method of threshold definition. Thresholds representing the UL of storms which did not produce debris flows represent a less conservative estimation of triggering conditions (e.g., higher rainfall intensities); however, debris flows are often initiated at lower rainfall intensities (Cannon et al. 2008). For example, a hypothetical database may contain occurrence and intensity information for 25 rainstorms, ten of which produced debris flows (Fig. 1). The LL) threshold model would define the threshold at the minimum rainfall intensity for the ten debris-flow events, regardless of the number of rainstorms that exceeded that intensity and did not produce a debris flow. The UL model would define the threshold at the maximum rainfall intensity of the 15 rainstorms which did not produce debris flows. Focusing on the 5-min duration in our hypothetical example (Fig. 1), the LL threshold is defined at an intensity of 13mm/h. This threshold results in 12 "false alarms", because 12 rainstorms had 5-minute rainfall intensities above the threshold but did not produce any debris flows. In contrast, the higher threshold of 25 mm/h defined by the UL model for the 5-min duration, does not have any false alarms. Instead, the UL threshold results in 4 "failed alarms", because 4 debris flows occurred during rainstorms with rainfall intensities below the threshold. An intermediate threshold between the UL and LL would provide a better balance between the number of false and failed alarms, and thus might be more appropriate for use in a warning system; however, there is not a yet a standard method to define such a threshold in a repeatable manner.

Here, we improve upon existing threshold models through the development of an objective set of regional rainfall ID thresholds for post-fire debris-flow initiation by (1) incorporating recent data on the precise timing of events relative to rainfall, (2) identifying the duration(s) at which ID thresholds provide the most accurate predictions of debris-flow occurrence, and (3) presenting an objective method to define the threshold intensities. We focus on thresholds developed from numerous close-proximity rain gauges for post-fire debris flows in the San Bernardino, San Gabriel and Santa Ana Mountains of southern California (Cannon et al. 2008, 2011), which initiated in response to high-intensity rainfall.

We use new data on the precise timing of 43 debris flow events to compare measures of peak storm rainfall intensity with the rainfall intensities that occurred prior to the recorded passage of debris flows. Data were collected from November 2007 to June 2008 on the Santiago fire (15 rainstorm records, one debris-flow event) and from October 2009 to June 2010 at the Station fire (244 rainstorm

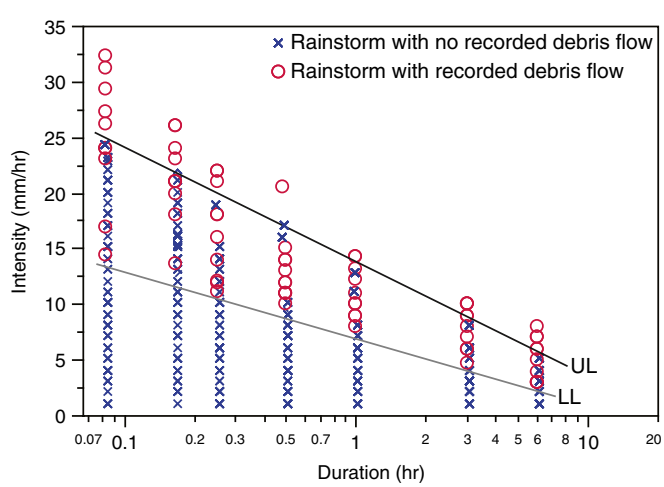

Fig. 1 Hypothetical dataset demonstrating the upper limit (*UL*) and lower limit (*LL*) ID threshold definition approaches. *Blue* ×'s represent rainfall intensities that did not produce debris flows. *Open red circles* indicated rainfall intensities that were associated with debris-flow events

records, 42 debris-flow events). This analysis lends insight into whether rainfall ID thresholds defined using peak storm intensities effectively characterize the rainfall conditions that initiate debris flows in recently burned watersheds. We then use receiver operating characteristics (ROC) analysis (Swets 1988; Fawcett 2006) to (1) identify the durations at which thresholds are the best predictors of debris-flow initiation and (2) objectively define critical threshold values for all analyzed durations. We objectively define thresholds using the Threat Score (TS) metric to maximize the number of correct predictions of debris-flow occurrence above ID threshold while minimizing the rate of Type I (false positive) and Type II (false negative) errors. Lastly, we compare the new, objectively defined threshold with the two thresholds defined by Cannon et al. (2008) for post-fire debris flow initiation in the same geographic region, as described below.

Existing regional thresholds

Cannon et al. (2008, 2011) defined a set of empirically derived rainfall ID thresholds for post-fire debris-flow initiation for two different geographic areas of southern California. The dataset presented in this paper was developed specifically for the steeper landscapes in the Transverse and Peninsular ranges, including the San Gabriel, San Bernardino, San Jacinto, Santa Ana and Palomar mountains (Fig. 2) (Cannon et al. 2008). The thresholds were defined using rainfall and debris-flow occurrence data collected in between 2003 and 2007. In all, the dataset used to define ID thresholds for this region consisted of peak rainfall intensity information from 81 rain gauge records from eight storms that produced a total of 82 debris flows within six burn areas. Cannon et al. (2008) calculated rainfall thresholds differently than most threshold studies. In most analyses, each point in the threshold plot (Fig. 3) represents individual rainstorms and landslide occurrences/non-occurrences. In the analysis of Cannon et al. (2008), each point represents the maximum rainfall intensity for each duration of measurement within a given storm and the corresponding debris flow response.

For the first year following wildfire, Cannon et al. (2008) defined the threshold for the UL of rainfall intensity–durations that produced debris flows by the equation $I=12.5D^{-0.4}$, where I is the rainfall intensity (mm/h) and D is the duration over which intensity was calculated (in hours). The lower limit (LL) of intensity–durations that initiated debris flow was represented by the equation $I=7.2D^{-0.4}$. Cannon et al. (2008) considered thresholds to be best defined at rainfall intensities measured over shorter rainfall durations (<20 h), as the post-fire debris flows analyzed here are not necessarily influenced by antecedent rainfall or soil moisture (Kean et al. 2011). The UL and LL models presented in Fig. 3 represent the thresholds developed by Cannon et al. (2008) and overlain on the data collected in the study areas described below.

Study areas

The current study analyzes data collected in 11 different drainage basins (Table 1). These basins are situated in two different burn areas located in the San Bernardino threshold region (Cannon et al. 2008): the Santiago (2007) and Station (2009) burn areas (Fig. 2). The Santiago fire burned 115 km^2 of steep terrain in the Santa Ana Mountains of Orange County. While multiple debris flows occurred during the winter of 2007–2008, precise eyewitness timing and rainfall data are available for only one debris-flow event in the Santiago burn area on 22 May 2008. The single analyzed basin in the Santiago burn area is underlain by the Jurassic Bedford Canyon formation and Cretaceous Santiago Peak volcanic (Morton 2004). The Bedford Canyon formation is characterized by slightly metamorphosed argillite, slate, greywacke, quartzite and limestone with lenses of conglomerate. The

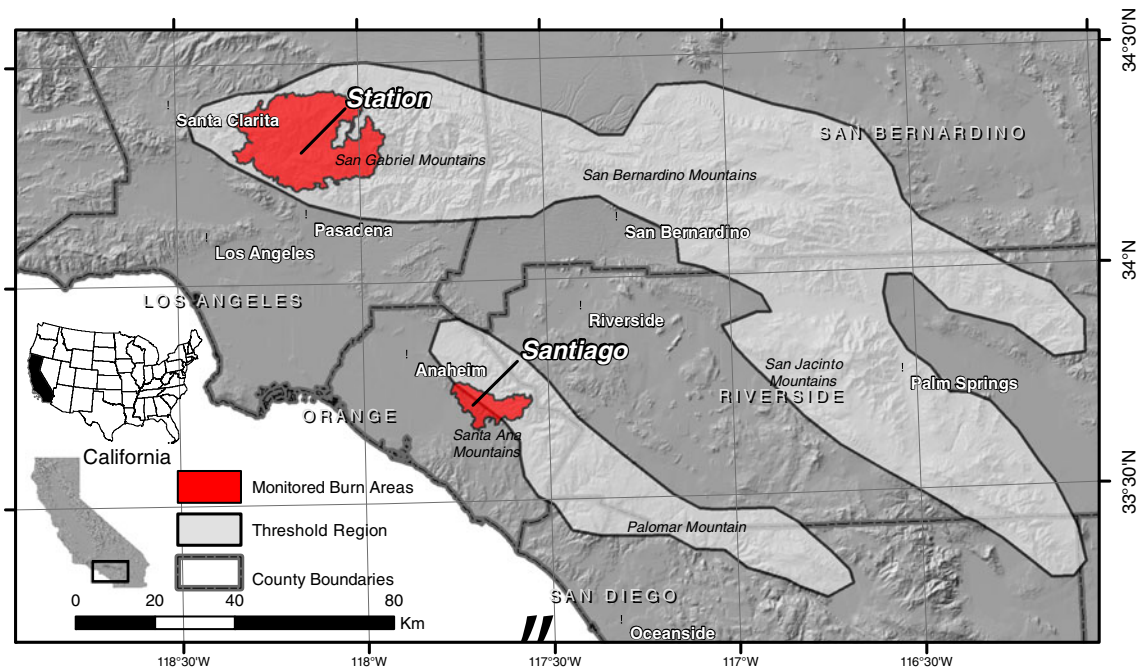

Fig. 2 Areas within the Transverse and Peninsular Ranges for which rainfall ID thresholds were defined by Cannon et al. (2008, 2010), and the locations of recent fires used to evaluate the effectiveness of the threshold method

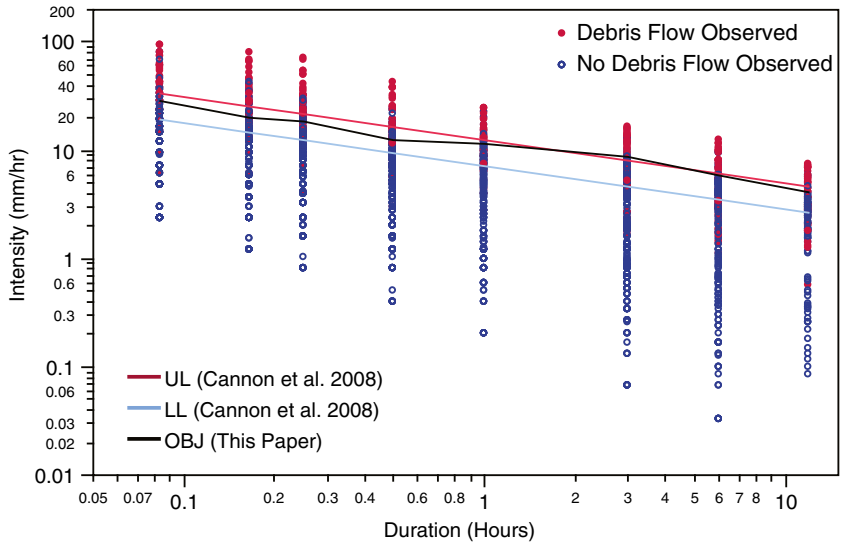

Fig. 3 Response data and objectively defined rainfall ID thresholds. The UL and LL models were described by Cannon et al. (2008) and were defined for a different dataset than that presented here. This dataset contains rainfall and initiation information for 15 rainstorms and one debris flow on the Santiago Fire (November 2008–June 2009) and 244 rainfall records and 42 recorded debris flows on the Station Fire (October 2009–June 2010). *Open blue circles* represent peak storm intensities for rainstorms that did not produce debris flows. *Red dots* represent triggering intensities for observed debris flows. *Black line* represents the objectively defined rainfall ID thresholds calculated for each duration. Note that each point represents the triggering rainfall intensity (if there was a debris flow) or peak rainstorm intensity (if there was no debris flow) at each duration of measurement for which storm duration was greater than measurement duration

Santiago Peak volcanics are comprised of hydrothermally altered basaltic andesite, andesite, dacite, rhyolite, breccia and welded tuff. There is evidence of abundant recent debris-flow activity near the study basin, although the exact dates of the events are not known.

The 650 km^2 station fire began on 26 August 2009 and was not fully contained until 16 October 2009. Given the size (the largest recorded in Los Angeles County) and potential debris-flow hazards (Cannon et al. 2010), the precipitation and hydrologic response of the area was intensely monitored during the winter of 2009–2010. Ten rain gauges installed in the burn area captured rainfall characteristics associated with five debris-flow-producing rainstorms (November 12–13, 2009, December 11–13, 2009, January 17–18, 2010, February 5–6, 2010, and February 27, 2010). The monitored basins are underlain by a highly faulted, complex mix of late Cretaceous granitic rocks comprised of quartz monzonites, granodiorites, tonalities, quartz diorites and diorites (Yerkes and Cambell 2005). Several of the study basins also experienced debris flows during the winters of 1928–1929, 1933–1934, 1968–1969, and 1977–1978 (Chawner 1935; Eaton 1935; Scott 1971; McPhee 1989; Cannon et al. 2011). There was evidence of more recent debris-flow activity in several of the basins within the Station burn area, although the exact date of the events are not known.

Data collection

In total, our database includes the rainfall conditions leading up to 43 debris-flow events and 259 storms that did not generate debris flows. To minimize the effects of decreases in debris-flow probability related to post-fire vegetation recovery, we only analyze data obtained during the first winter following wildfire (which typically occur in the mid- to late Fall). To ensure that the analyzed areas were at least minimally susceptible to debris flow, we only included sites that had at least one debris-flow producing rainstorm during the 1-year period of record. In addition, we considered only records from rain gauges located within 2 km of a debris-flow producing watershed to minimize the effects of the spatial variability of rainfall amounts and intensity. Our dataset represents the most comprehensive known database of post-fire debris-flow initiation conditions and event timing. The timing of all but one event in the database was recorded by channel monitoring equipment installed at a stable cross-section in 12 different watersheds. Timing information for the single event not recorded by a monitoring site was obtained from timestamps on an eyewitness' camera, and rainfall data were obtained from a nearby (~400 m planimetric distance) rain gauge installed by the authors.

For all sites, rainfall amounts were recorded by Onset RG3-M tipping-bucket rain gauges using Hobo data loggers.[1] For all but one rain gauge, a timestamp is recorded at every 0.2 mm of rainfall accumulation. The exception was located near the Gould study basin, where accumulation data were binned into 5-min intervals to meet telemetry requirements.

Channel monitoring equipment included non-contact stage gauges (Kean et al. 2011) or pore-pressure transducers (Kean et al. 2012). Non-contact stage gauges, mounted on a bridge above the channel bed, include both laser and ultrasonic distance meters that measure the height (stage) of channelized flow at a constant sampling rate (10 Hz for the laser sensors, 2 s for the sonic sensors) during rainfall. While the sampling interval differs, both instruments accurately measure flow stage to within ~1 cm. As shown in Fig. 4, the passing front of a debris-flow surge is easily identifiable as a rapid rise in stage with a relatively long-lived recessional flow (McCoy et al. 2010; Kean et al. 2011, 2012).

Pore-pressure transducers installed in bedrock in the channel bed measured the fluid pressure head (recorded at 1-min intervals) during a flow event. Internal memory limitations within the pore

[1] Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Table 1 Summary data of drainage basins included in this study

Site	Burn area	Year burned	Range	Debris flow timing source	No. of debris flow events	Easting (UTM NAD 83, Zone 11, m)	Northing (UTM NAD 83, Zone 11, m)	Basin area (km^2)	Maximum basin length from outlet to divide (m)
Modjeska Canyon	Santiago	2007	Santa Ana	Eyewitness photo	1	442,746	3,730,279	0.543	1,801
Arroyo Seco	Station	2009	San Gabriel	Laser	13	389,956	3,788,964	0.135	206
Big Tujunga	Station	2009	San Gabriel	Laser	2	386,462	3,794,688	1.370	3,030
Dunsmore 1	Station	2009	San Gabriel	Laser	6	385,649	3,791,625	0.475	1,415
Dunsmore 2	Station	2009	San Gabriel	Sonic	3	385,225	3,790,898	0.795	734
Gould	Station	2009	San Gabriel	Pressure transducer	5	390,058	3,788,019	0.328	908
Mullaly	Station	2009	San Gabriel	Pressure transducer	4	387,744	3,789,635	0.628	1,880
Oak Creek	Station	2009	San Gabriel	Pressure transducer	1	385,448	3,790,460	0.019	225
Shields Canyon	Station	2009	San Gabriel	Pressure transducer	5	386,075	3,790,540	0.363	1,356
Starfall	Station	2009	San Gabriel	Pressure transducer	2	386,217	3,790,279	0.286	1,290
Winery	Station	2009	San Gabriel	Pressure transducer	1	389,060	3,788,352	0.183	927

pressure transducer prevented higher sampling frequencies. Despite the limitations related to sampling interval, pore pressure transducers have proven to be effective at resolving the timing of debris-flow passage to within a few minutes (Kean et al. 2012).

The measurements made at the monitoring sites (and timestamps provided by eyewitness photographs) provide accurate (±3 min) data regarding the timing of debris-flow passage at the instrument location. In order to relate the rainfall conditions that triggered the debris-flow event to the recorded time of passage, we compared rainfall data collected either at the monitoring site or at a nearby rain gauge. In both cases rainfall timestamps have been synchronized with the times recorded at the stage gauges and pressure transducers. We consider the highest rainfall intensities (measured at each duration) recorded within the 30-min period prior to the recorded passage of debris flow to be the triggering rainfall intensity. Given the small area and relatively short maximum flow lengths in the analyzed basins (Table 1), it is unlikely that rainfall rates at higher elevations in the drainage basin deviated substantially from recorded values. In addition, it is unlikely that debris flows initiated more than 30 min prior to the recorded time of debris flow, because of the small size of the drainage basin, published velocities of natural debris flows (Berti et al. 1999; Arattano and Marchi 2000, 2005; Tecca and Genevois 2009; Tang et al. 2011; Okano et al. 2012; Yong et al. 2012), and close correlation between 5-min rainfall intensity and flow stage in the study basins (Kean et al. 2011). Henceforth, the time at which the highest rain rates within the 30 min preceding debris-flow passage were recorded will be referred to as the "triggering" time, and the corresponding rainfall intensity will be referred to as the "triggering" intensity.

Analysis methods

In this section, we describe the method by which we calculate rainfall intensities. We then provide an overview of ROC methods and describe how we use the ROC utility functions of true positive rate (tp_{rate}), false positive rate (fp_{rate}), area under ROC curve (AUC) and TS to optimize and evaluate critical values for rainfall ID thresholds.

Rainfall analysis

From the data collected by the rain gauges, we calculate two classes of rainfall intensities: peak storm intensity and triggering intensity (Fig. 4). Rainfall intensities were calculated for every minute of all rainstorms by backwards differencing cumulative rainfall over durations of 5, 10, 15, 30, 60, 180, 360, and 720 min, as follows:

$$I_D(t) = \frac{(R(t) - R(t - D))}{D}$$

where I_D is the rainfall intensity (mm/h), t is time, R is the cumulative rainfall, and D is the selected duration (in hours).

Peak storm intensity is the maximum rainfall intensity recorded during the storm at each of the analyzed durations. For each storm on record, we identify the peak storm intensity, the time at which peak storm intensity occurred, and the presence or absence of a debris flow during that rainstorm. For storms that did not produce debris flows, we identified the peak rainstorm intensities and classified them as non-triggering rainfall intensities.

For rainstorms containing multiple debris-flow events (i.e., multiple distinct sets of debris flow surges), we document the

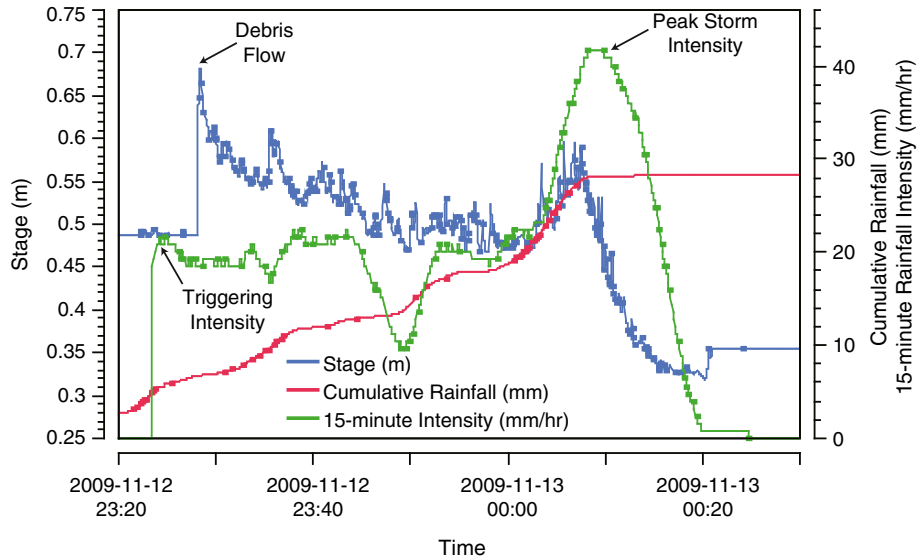

Fig. 4 Graph illustrating the measures of cumulative rainfall, 15-min intensity, flow stage and the difference between triggering intensity and peak storm intensity. The 15-min triggering intensity is less than the 15-min peak storm intensity, and the 15-min peak storm intensity occurred after the occurrence of the debris flow. Record from the November 12–13, 2009 storm and debris flow in the Arroyo Seco monitoring basin, Station fire

triggering rainfall intensity for each distinct period of debris flow, which are separated by at least 20 min with little or no measureable channel flow. Given the high correlation between flow stage and 5-min rainfall intensity (Kean et al. 2011), published debris flow velocities, and the flashy response of the hydrograph to rainfall, we feel that 20 min is adequate for defining individual debris-flow surges.

We selected an 8-h period without rainfall as the minimum inter-storm period. This value helps discriminate the storm durations common during the southern California winter, which are characterized by short-duration, high-intensity convective thunderstorms and long-duration low-intensity frontal storms. The 8-h inter-storm period is able to reliably differentiate both types of rainstorms. A longer inter-storm period (e.g., 24 h) may not adequately separate two distinct convective thunderstorms, whereas a shorter inter-storm period (e.g., 4 h) may unnecessarily split a longer duration frontal storm into separate storm events.

Receiver operating characteristic analysis

ROC analysis is a technique that is used for assessing the performance of classification models (Swets 1988; Fawcett 2006). ROC analysis can be used to evaluate models that produce discrete (two or more classes) or continuous (e.g., probability) predictions. Recently, ROC has been used to assess the performance of landslide susceptibility models (Carrara et al. 2008; Godt et al. 2008; Frattini et al. 2009; Baum et al. 2010; Cervi et al. 2010). Here, we use ROC analysis to both evaluate the performance of the Cannon et al. (2008) rainfall ID thresholds for southern California and to objectively define a new threshold based on triggering rainfall intensities defined using the additional data regarding debris-flow timing.

Rainfall ID thresholds represent a binary classifier model where there are four possible outcomes which can be graphically represented using a contingency table (Fig. 5a). Event occurrence is considered to be either True or False (it either occurred or did not occur), while model predictions are considered to be Positive or Negative (successful prediction or prediction failure). The relationship between model prediction and event occurrence is then assigned to one of four classes. A true positive (TP) represents an event where rainfall rates exceeded threshold, and a debris flow was recorded. A true negative (TN) represents an event where rainfall rates were below critical threshold and no debris flows were recorded. False positive (FP) events occur when rainfall rates exceed threshold but no debris flows were triggered. This can also be considered a "false alarm" or type I error. A false negative (FN) occurs when rainfall rates were below threshold, yet a debris flow was recorded. This represents a "failed alarm" situation or Type II error.

ROC utility functions (Fig. 5a) are used to quantitatively describe the performance of the each threshold model by assessing the distribution of outcomes within the four classes. Basic ROC utility functions include the true positive rate and false positive rate. The true positive rate (tp_{rate}, also referred to as the hit rate, sensitivity or recall) measures how well the model predicts the occurrence of an event. In this study, the true positive rate measured the proportion of all debris flows that occurred when rainfall rates exceeded threshold. The false positive rate (fp_{rate}), measures the rate at which the model predicts a positive outcome, yet the event does not occur. In this study, fp_{rate} represents the proportion of rainstorms that had no observed debris flows yet still exceeded threshold value. The relationship between tp_{rate} and fp_{rate}, when plotted in ROC space (Fig. 5b), is used as a relatively simple means of graphically depicting the predictive capability of the classification model (Swets 1988; Fawcett 2006). The relationship between tp_{rate} and fp_{rate}, when plotted in ROC space, is also used to calculate the AUC (Fig. 5b). AUC values may be used to assess the relative value of individual classifiers, where higher AUC values represent better prediction models (Fawcett 2006) (Fig. 5b). In this study, we utilize AUC to identify the measurement durations that are the best predictors of debris flow initiation.

Objective threshold definition

We quantitatively assess the predictive ability of the objectively defined thresholds at durations from 5 to 720 min from triggering intensities using the TS metric. TS represents a measure of the overall performance of the classifier model where a perfect model score would equal one, and each incorrect prediction (FP or FN) reduces the value of TS (Fig. 5a) (Schaefer 1990). We chose to use

A)

MODELED CLASS	OBSERVED CLASS: Debris Flow	OBSERVED CLASS: No Debris Flow
Above Threshold	**True Positive (TP)** *Above Threshold, Debris Flow Observed*	**False Positive (FP)** *Above Threshold, No Debris Flow Observed*
Below Threshold	**False Negative (FN)** *Below Threshold, Debris Flow Observed*	**True Negative (TN)** *Below Threshold, No Debris Flow Observed*

$$TP_{rate} = \frac{TP}{TP + FN}$$

$$FP_{rate} = \frac{FP}{FP + TN}$$

$$TS = \frac{TP}{TP + FN + FP}$$

B)

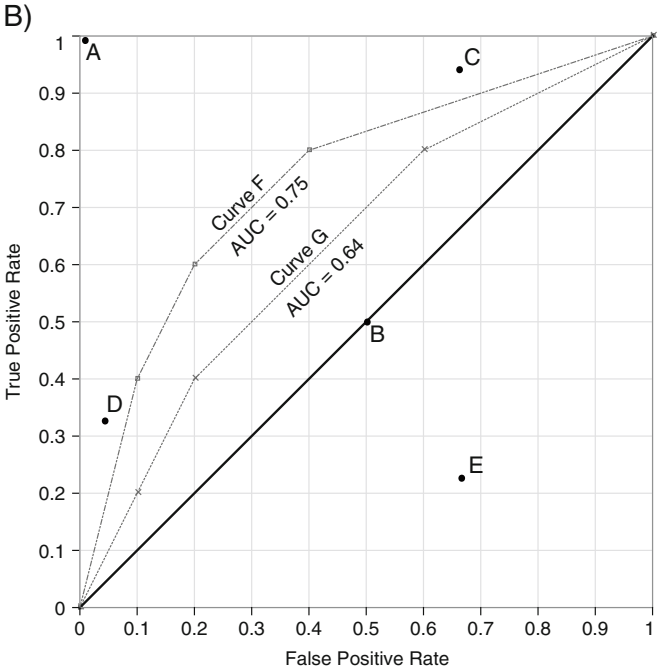

Fig. 5 ROC analysis methods and metrics. **a** 2×2 contingency table identifying the four possible outcomes of a binary classifier model and common measurements of classifier performance used in ROC analysis. **b** ROC space, hypothetical model results and AUC curves. Point *A* represents a perfect model, *B* represents a model no better than a random guess, point *C* represents a model that correctly predicts a positive outcome at the expense of a high false-alarm rate, point *D* represents a model that minimizes the false alarm rate at the expense of correctly predicting positive outcomes, point E represents a model where the prediction is worse than a random guess. Curves *F* and *G* represent hypothetical ROC curves with associated AUC values, with curve *F* representing a better-than-average performance

the TS as the metric used in the optimization analysis, because it equally weights the reduction in score for both FN and FP events, and is an easily calculated measure that can be updated during a storm season for areas where existing thresholds are unavailable.

We used TS to define the ID thresholds by iteratively calculating TS for all analyzed durations for intensities ranging from 1 to 100 mm/h at 0.1-mm/h intervals. The threshold was defined at the maximum value of TS (Fig. 6). The highest rainfall intensity was defined as the critical ID threshold when maximum TS value was calculated for multiple rainfall intensities. The objectively defined threshold model will subsequently be referred to as the OBJ model.

Results and discussion

Peak storm intensity, triggering intensity and debris-flow timing

In many rainfall ID threshold analyses, the exact timing of debris-flow occurrence is typically unknown (Guzzetti et al. 2007). In lieu of such information, it is commonly assumed that measures of peak storm rainfall intensity are representative of the conditions that trigger debris flow. Additionally, it is assumed that the peak rainstorm intensities occur approximately at the time of debris-flow initiation. We test these assumptions for 43 post-fire debris flows in southern California by comparing (1) peak rainstorm intensities to triggering rainfall intensities and (2) the time at which peak rainstorm intensities were measured to the recorded time of debris flow.

We calculated the statistical difference between peak storm intensities and triggering intensities (Fig. 7). We test for statistically significant difference using non-parametric Wilcoxon signed-rank test (Wilcoxon 1945). For this test, χ^2 scores greater than or equal to the $\chi_{critical}^2$ (in our case, $\chi_{critical}^2=3.84$) indicated statistically significant differences between samples (in our case, triggering intensities and peak storm intensities). This test was used because the sample populations were not normal and did not have equal variances. Statistically significant differences ($\alpha=0.05$) were identified at all analyzed durations between peak storm intensities and those that occur immediately prior to a recorded debris flow. For all durations, the triggering rainfall intensities were significantly lower than peak storm intensities (Fig. 7). χ^2 scores and p values (the probability that χ^2 score$\geq\chi_{critical}^2$, where $\alpha=0.05$, significance is obtained when $p<\alpha$) are reported in Fig. 7.

We also analyzed the differences between times of peak rainstorm intensities and the time at which a debris flow was recorded at a monitoring location (Table 2). In general, peak rainstorm intensities occurred after the recorded time of a debris flow surge, and the average time difference between peak storm intensity and debris flow occurrence increased for increasing rainfall durations (Fig. 8). Rainfall intensities measured over a 5-min duration yielded the lowest mean time differences, with peak-storm intensities occurring, on average, 35 min after a recorded debris flow. Rainfall intensities measured over 720 min displayed the greatest time differential, with peak-storm intensities occurring, on average, more than 3 h after recorded debris flows. In addition, the peak storm intensity at the 720-min duration occurred after the debris flow in over 94 % of the debris flow events. It is not unexpected that post-fire debris flows are initiated at rainstorm intensities lower than peak values. However, as will be shown in the next section, an important consequence of these results is that rainfall thresholds derived from peak rainfall intensities can have a higher rate of failed alarms than thresholds based on triggering intensities. Our analysis quantifies the error associated with the location of ID thresholds at different rainfall intensities, and provides an objective method for defining ID thresholds based upon highly precise and accurate debris-flow initiation data.

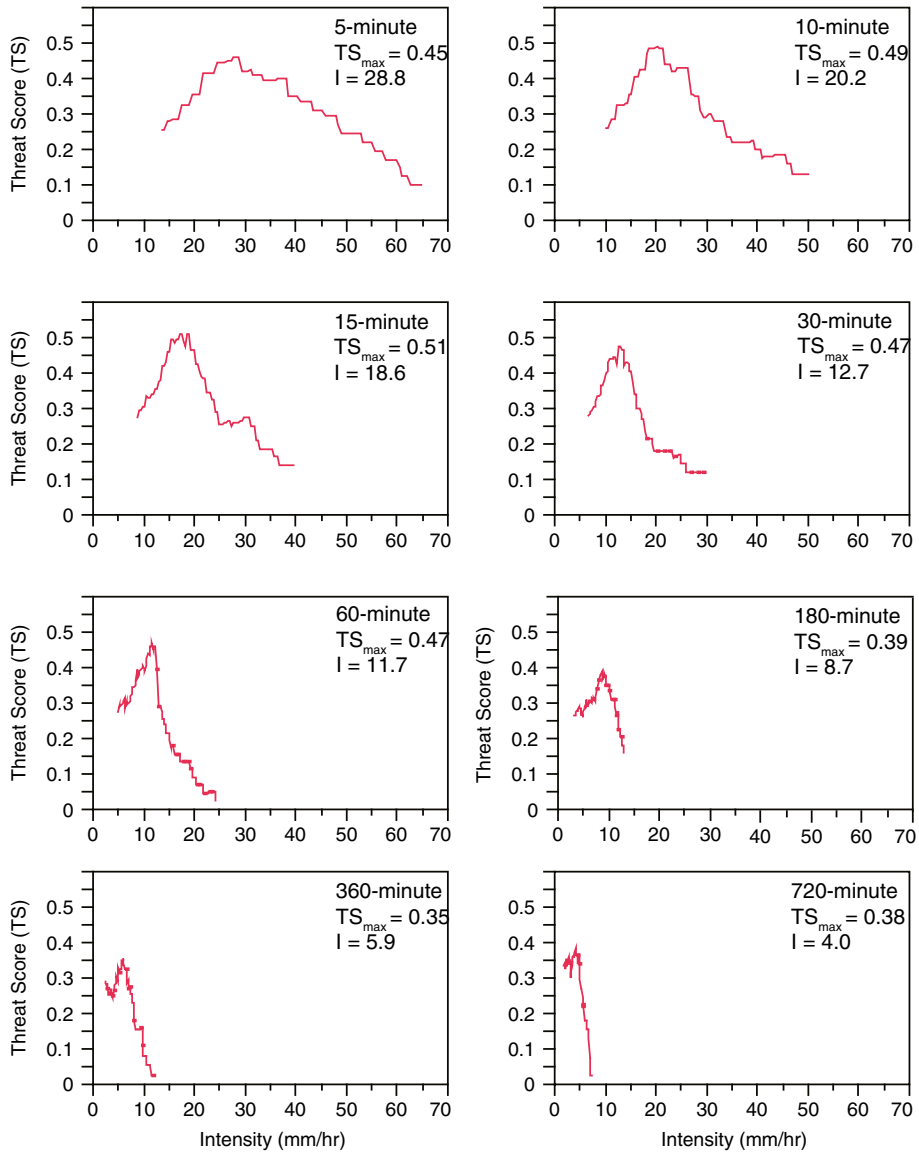

Fig. 6 Graph showing the relationship between a critical rainfall threshold established at a given rainfall intensity and threat score (*TS*). Thresholds were defined at the intensity where TS was maximized (TS_{max})

Objective rainfall intensity–duration thresholds from triggering intensity data

Given the statistically significant differences between peak storm intensities and triggering intensities, and the large time difference between peak storm intensity and the time of a recorded debris-flow surge, we use the triggering rainfall intensity data to objectively define a new set of critical rainfall ID thresholds for the initiation of post-fire debris flows (Fig. 3). We objectively defined the rainfall ID thresholds at durations of 5, 10, 15, 30, 60, 180, 360, and 720 min by maximizing the utility function of TS (Fig. 6, Table 3).

We chose to define and analyze critical rainfall intensity thresholds at individual durations rather than as a power law in order to maximize the predictive ability of the threshold at each analyzed duration. However, our thresholds resemble a power law given by the equation: $I=12.4D^{-0.4}$ ($r^2=0.75$).

ROC analysis of thresholds

ROC analysis was performed on the triggering rainfall intensities for recorded debris flows, and for the peak storm intensities for rainstorms where no debris flows were produced. The Cannon et al. (2008) UL and LL thresholds were defined from measurements of peak storm rainfall intensity, while the objective thresholds (OBJ) were defined from triggering rainfall intensities using the methods described in the previous section.

For all analyzed models, the best predictions of debris flow occurrence were identified at durations less than or equal to 30 min as demonstrated by AUC values (Fig. 9). This result is consistent with direct measurements of runoff-generated debris-flow timing relative to rainfall (Kean et al. 2011; Okano et al. 2012). Okano et al. (2012) identified 10-min and 24-h durations as the best predictors for debris-flow occurrence on the slopes of Mount Yakedake in Japan, while Kean et al. (2011) found highest correlations between rainfall intensity and peak flow-stage at durations

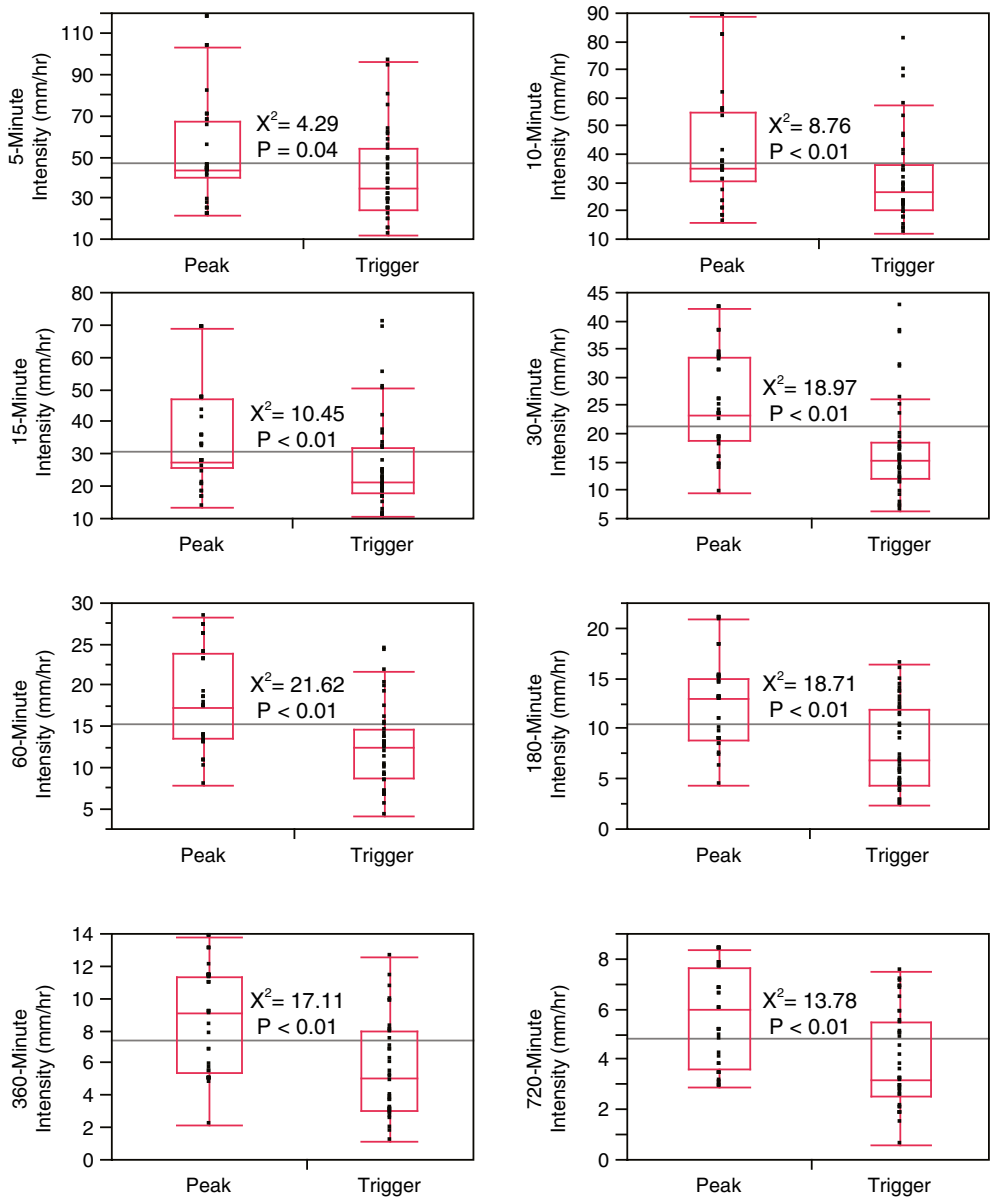

Fig. 7 Box and whisker plot representing the differences between peak rainstorm intensities and triggering intensities. *Black dots* represent data points. *Upper box end* represents 75th percentile, *Lower box end* represents 25th percentile. Group mean is represented by the *horizontal red line inside the rectangle*. Whiskers represent 90th and 10th percentiles. *Horizontal black line* represents the grand mean (mean of both peak intensity and triggering intensity values)

Table 2 Descriptive statistics for the amount of time difference between observed debris flows and peak storm intensity

Duration (min)	Mean difference (min)	Median difference (min)	% Peak intensity occurring after debris flow
5	35.4	2.9	57.1 %
10	56.2	2.9	61.5 %
15	47.4	5.0	59.2 %
30	68.5	12.0	69.6 %
60	75.4	36.0	72.1 %
180	56.6	76.5	76.2 %
360	104.0	145.0	89.5 %
720	187.7	157.0	94.6 %

Positive values indicate peak storm intensity occurring after a debris flow.

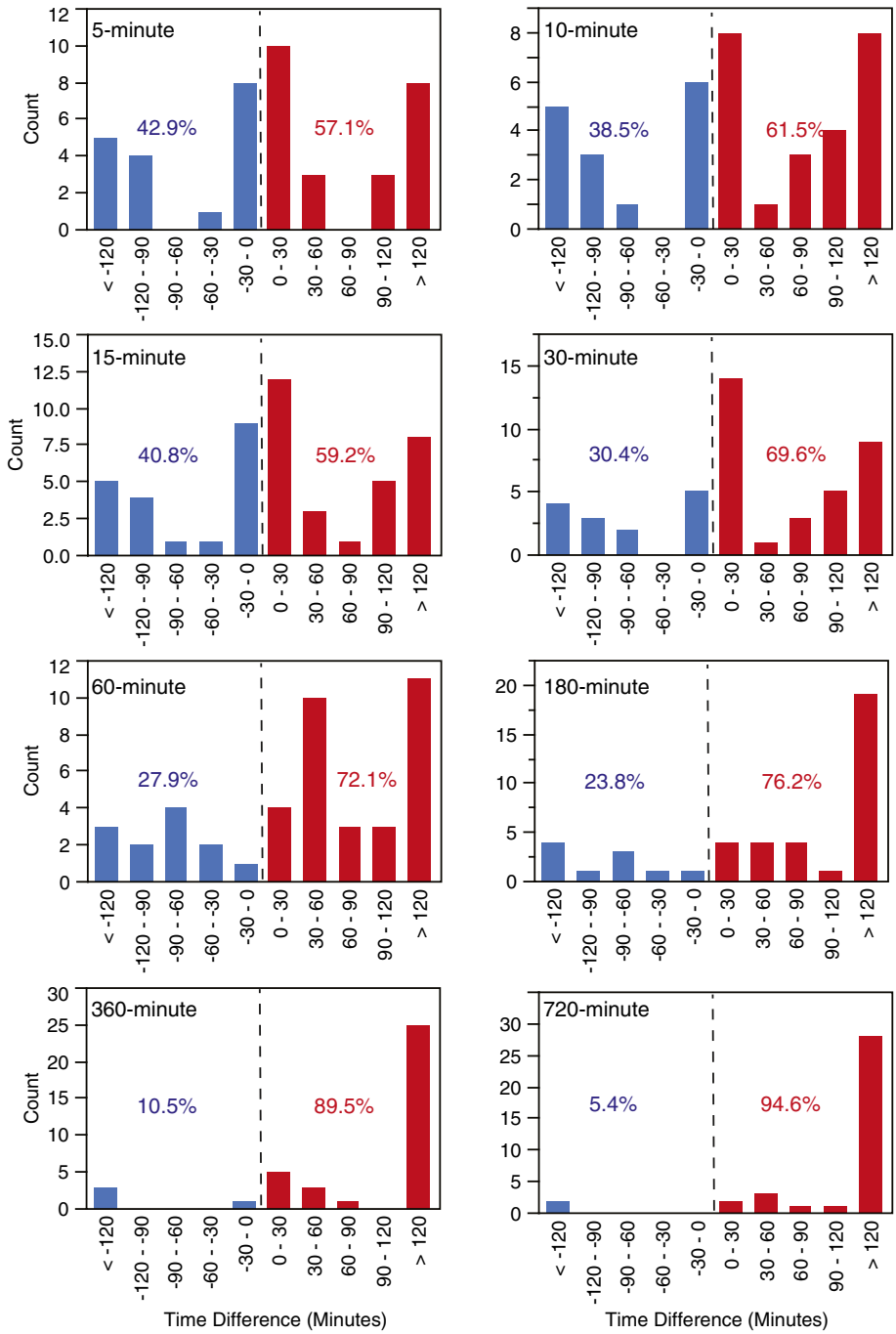

Fig. 8 Histograms showing the distributions of differences between recorded time of debris flow and the peak storm intensities of varying durations. Positive values (*red*) indicate peak storm intensities occurring after the debris flow; negative values (*blue*) indicate peak storm intensities occurring prior to the debris flow. Triggering intensities are situated within the −30 to 0 bin

≤30 min for the same recently burned watersheds analyzed in this study. Kean et al. (2011) attributed this correlation to the mechanics by which the post-fire debris flows initiated. Long-duration (>1 h) thresholds are typically used to model systems where rainfall infiltration produces decreases in slope stability, resulting in slope failure by landsliding and the transition of the slope failure to debris flow (Caine 1980). Rather than infiltration contributing to landsliding and subsequent transition to debris flow, post-fire debris flows are considered to be often initiated from surface-runoff processes. The transition from water flow to debris-flow may occur by means of progressive sediment bulking (Cannon et al. 2001; Cannon and Reneau 2001; Cannon et al. 2003), failure of the channel bed (Takahashi 1981), the entrainment of material by a concentrated flow of clear water (commonly referred to as the "firehose" effect) (Johnson and Rodine 1984), decimeter-scale Coloumb failures (Schmidt et al. 2011) or some combination of processes. As our best predictors of debris-flow occurrence were associated with durations ≤30 min, our results support these findings, and further emphasize the need to better understand the physical processes that allow for the transition from infiltration-excess overland flow to debris flow in recently burned watersheds.

The tp_{rate} and fp_{rate} derived from both the peak storm intensity dataset and the triggering intensity dataset were plotted in

Table 3 Rainfall ID thresholds and associated ROC metrics for each threshold model

Duration (min)	Threshold (mm/h)	n	TN	FN	FP	TP	tp_{rate}	fp_{rate}	TS
Cannon et al. (2008) upper limit (UL) threshold model									
5	33.8	254	206	19	10	19	0.50	0.05	0.40
10	25.6	254	205	17	11	21	0.55	0.05	0.43
15	21.8	259	207	23	9	20	0.47	0.04	0.38
30	16.5	256	206	26	9	15	0.37	0.04	0.30
60	12.5	255	207	23	6	19	0.45	0.03	0.40
180	8.1	241	189	23	10	19	0.45	0.05	0.37
360	6.1	179	135	22	7	15	0.41	0.05	0.34
720	4.6	134	93	24	2	15	0.38	0.02	0.37
Cannon et al. (2008) lower limit (LL) threshold model									
5	19.5	254	161	5	55	33	0.87	0.25	0.35
10	14.7	254	158	4	58	34	0.89	0.27	0.35
15	12.5	259	155	6	61	37	0.86	0.28	0.36
30	9.5	256	157	5	58	36	0.88	0.27	0.36
60	7.2	255	143	8	70	34	0.81	0.33	0.30
180	4.6	241	126	11	73	31	0.74	0.37	0.27
360	3.5	179	84	12	58	25	0.68	0.41	0.26
720	2.7	134	56	12	39	27	0.69	0.41	0.35
Objectively defined (OBJ) threshold model									
5	28.8	254	195	11	21	27	0.71	0.10	0.46
10	20.2	254	195	9	21	29	0.76	0.10	0.49
15	18.6	259	202	14	14	29	0.67	0.06	0.51
30	12.7	256	192	11	23	30	0.73	0.11	0.47
60	11.7	255	204	18	9	24	0.57	0.04	0.47
180	8.7	241	192	23	7	19	0.45	0.04	0.39
360	5.9	179	134	21	8	16	0.43	0.06	0.36
720	4.1	134	87	22	8	17	0.44	0.08	0.38

ROC space to compare the performance of each prediction type (Fig. 10). In this study, tp_{rate} represents the fraction of debris flows that were recorded at rainfall intensities above threshold and where no debris flows were recorded at intensities below threshold. The false positive rate represents the proportion of rainstorms that had no observed debris flows yet still exceeded critical rainfall intensities (false alarm or type I error). Visual inspection of the ROC graph suggests the best predictions are obtained by the UL and OBJ models at durations ≤30 min (Fig. 10), as these points fall closest to the upper left corner of the graph ($x=0$, $y=1$) where a perfect classifier would be found (point A on Fig. 5b). OBJ thresholds at these durations had higher tp_{rate} values than UL thresholds, at the cost of higher slightly higher fp_{rate} values.

The LL thresholds were found to have both the highest tp_{rate} and fp_{rate} values, while the UL thresholds were found to have the lowest tp_{rate} and fp_{rate} values. The OBJ threshold model represented an intermediary between the UL and LL models for both tp_{rate} and fp_{rate}. For all models, the highest tp_{rate} values were identified for rainfall intensities measured over shorter durations. The tp_{rate} values for the LL model ranged from 0.68 to 0.89, with the highest values found for durations ≤30 min and the lowest values found for durations ≥360 min. UL model tp_{rate} values ranged from 0.38 to 0.55, with highest values evident at the 5- and 10-min durations, and the lowest values at durations of 30, 360, and 720 min. OBJ model tp_{rate} values ranged from 0.43 to 0.76, with the highest values identified for durations ≤30 min.

For all models, fp_{rate} values tended to have smaller ranges than those for tp_{rate}. Overall, the lowest fp_{rate} values were identified for the UL model, and the highest identified for the LL model. UL fp_{rate} values ranged from 0.02 to 0.05, with the lowest values identified at durations of 60 and 720 min. LL fp_{rate} values ranged from 0.25 to 0.41, with the lowest values identified at durations ≤30 min. In theOBJ model, these values were constrained to a range between 0.04 and 0.11.

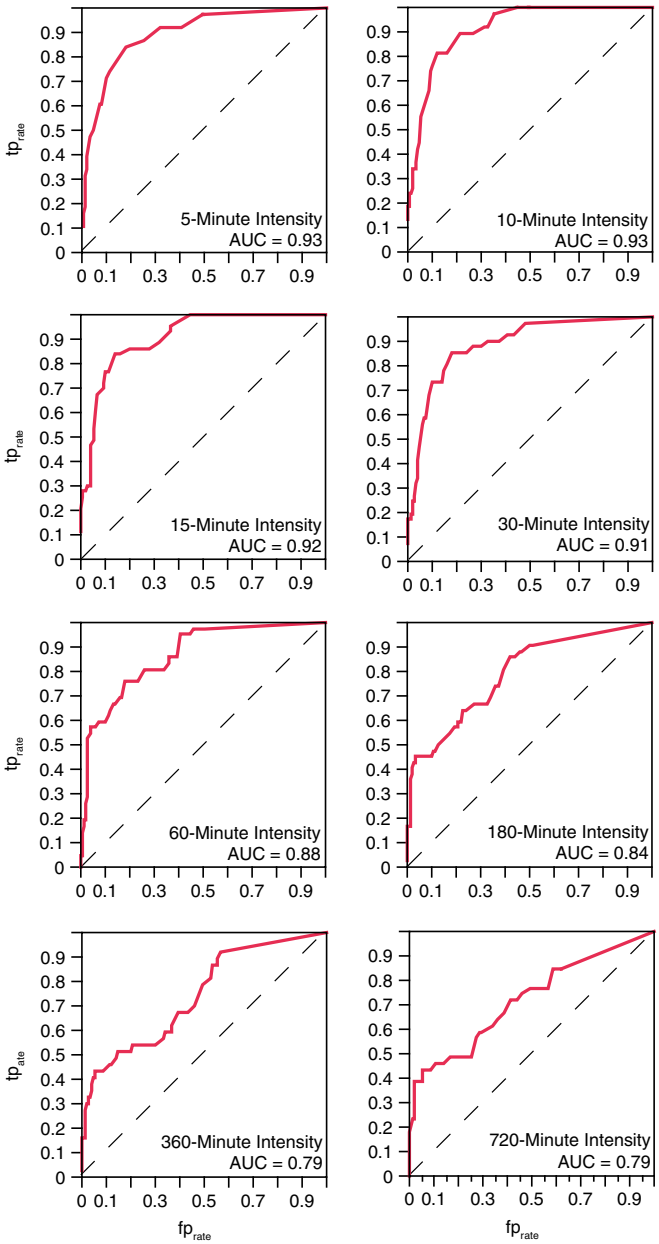

Fig. 9 Area under ROC curve (*AUC*) graphs for rainfall and response data at each duration of measurement

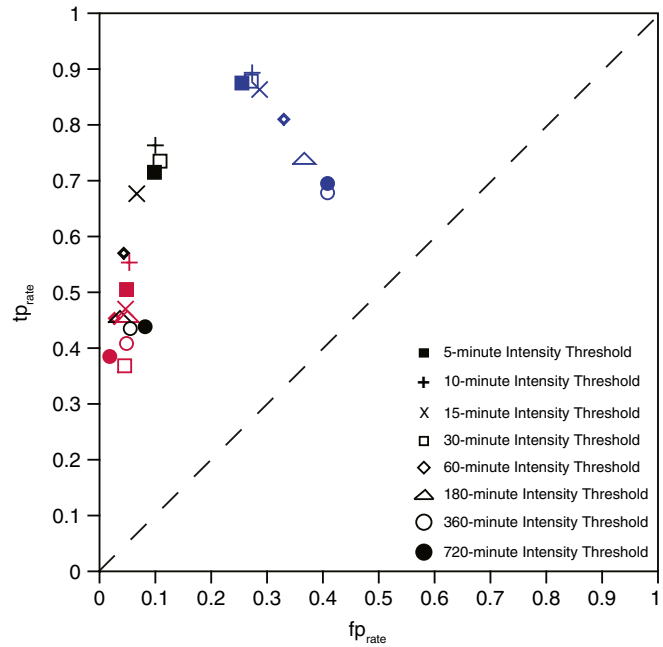

Fig. 10 ROC graph for thresholds at each duration, calculated for LL (*blue*), UL (*red*), and OBJ thresholds from triggering intensities (*black*)

OBJ values ranged from 0.26 to 0.36 for the LL model, with highest OBJ values identified at durations ≤60 min. For the UL model, OBJ values ranged from 0.30 to 0.43, with the highest values obtained at the 5-, 10-, and 60-min durations. Optimizing the OBJ produced OBJ values ranging from 0.36 to 0.51, with the highest values identified at durations ≤60 min (Fig. 6).

The objectively defined thresholds provide the best balance between correct predictions and low false alarms. The short (≤30 min) duration tp_{rate} values were an average of 15 % greater than those based upon the UL model, and 21 % less than those for the LL model. When comparing the OBJ to LL and UL models, the reduction in tp_{rate} compared to the LL model is offset by a substantial 25 % reduction in fp_{rate}, and only a modest increase of 3 % in fp rate relative to the UL threshold model. OBJ values were an average of 7 % greater than those based upon the UL model, and 11 % greater than those for the LL model.

Threshold exceedance and the timing of debris flows

Here, we assume that within a range of rainfall intensities at a given duration, there is a point where a slight increase in rainfall intensity yields a dramatic increase in the probability of debris-flow occurrence. As such, we would expect that the critical rainfall intensity would be exceeded at a point during a rainstorm immediately prior to the recorded time of debris flow. To test the ability of a critical rainfall intensity threshold to predict the within-storm occurrence of a debris flow, we analyzed the difference between the time at which a debris flow was recorded and the time at which rainfall rates exceeded threshold (Fig. 11). Descriptive statistics of the mean, median, % of debris flows that were recorded within 30 min of threshold exceedance, and % of debris flows occurring at some point in the storm after thresholds were exceeded were used to assess the predictive capability of each model at different measurement durations, and are summarized in Table 4. We consider a threshold to be a good predictor of debris flow occurrence if the critical rainfall intensity threshold is exceeded prior to the recorded time of a debris-flow surge, represented as negative values in Fig. 11 and Table 4.

For the UL model, the exact timing of recorded debris flows was not consistently predicted at any duration. The greatest proportion (33 %) of debris flows occurring within 30 min of threshold exceedance was identified at rainfall intensities measured for durations of 30 min. For all durations, threshold exceedance at some point during the rainstorm prior a recorded debris flow never exceeded 44 % (D=30 min).

The LL model performed significantly better than the UL model, with a majority of surges occurring within 30 min of threshold exceedance for rainfall intensities measured at

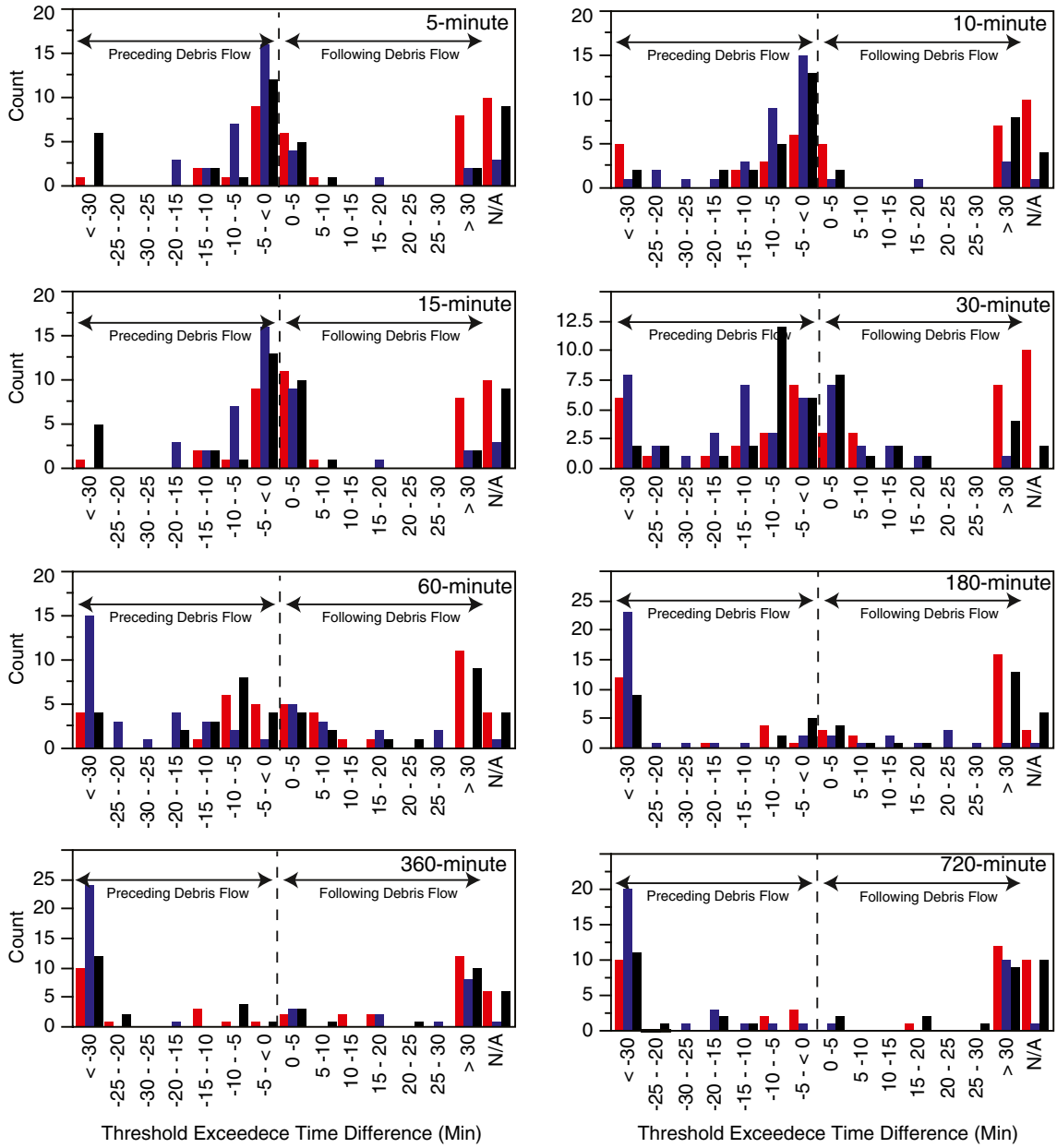

Fig. 11 Histograms illustrating the difference between the time at which critical rainfall threshold intensities were exceeded and the recorded time of debris flow for LL (*blue*), UL (*red*) and OBJ (*black*). Negative values indicate thresholds were exceeded prior to debris flow, positive values indicate that thresholds were exceeded after debris flows

durations ≤15 min. For all measured rainfall durations, thresholds were exceeded at some point prior to a debris flow for a majority of all recorded events. However, the ability of the LL model to predict the within-storm timing of debris flows is tempered by the high false alarm rate identified in the previous sections.

The OBJ model performed better than the UL model but not as well as the LL model in predicting the within-storm timing of debris flows. A majority of debris flows occurred within 30 min of threshold exceedance for the 30-min threshold. For both the 10- and 30-min ID thresholds, more than 50 % of debris flows occurred after the ID threshold was exceeded.

While all models proved to be limited in predicting the exact time of debris flow occurrence, the OBJ model represents an improvement over the UL model. In addition, we feel that the dramatic improvements in fp_{rate} compared to the LL model counteract the reduction in the difference between exceedance times and debris flow occurrence identified in the OBJ model.

The lack of complete correspondence between the time at which thresholds are exceeded and the timing of debris flow initiation for all threshold models illustrates the limitation of the threshold approach for predicting the exact time of post-fire debris flow occurrence. Rainfall thresholds derived from triggering rainfall intensities (our OBJ thresholds) are only able to correctly predict debris-flow occurrence a maximum of 50 % of the time. Improved prediction rates will require a better understanding of the mechanisms of runoff-induced debris-flow initiation in recently burned watersheds, which are

Table 4 Summary of differences between recorded time of debris flow and the time at which thresholds were exceeded

Duration (min)	Mean difference (min)	Median difference (min)	% of Debris flows where thresholds were exceeded ≤30 min prior to recorded surge	% of debris flows that occurred after threshold was exceeded
Cannon et al. (2008) upper limit (UL) threshold model				
5	36.7	0	27.9 %	28.3 %
10	−7.3	−1.5	25.6 %	34.8 %
15	36.7	0	27.9 %	28.3 %
30	−0.2	−2.4	32.6 %	43.5 %
60	17.5	0.5	27.9 %	34.8 %
180	−4.8	2.0	14.0 %	39.1 %
360	−7.5	0.0	14.0 %	34.8 %
720	−12.0	−3.1	11.6 %	32.6 %
Cannon et al. (2008) lower limit (LL) threshold model				
5	1.7	−3.0	60.9 %	60.9 %
10	−2.2	−4.0	67.4 %	67.4 %
15	1.7	−3.0	60.9 %	60.9 %
30	−20.1	−7.6	47.8 %	65.2 %
60	−40.5	−16.6	30.4 %	63.0 %
180	−81.4	−80.0	13.0 %	63.0 %
360	−57.4	−57.0	2.2 %	54.3 %
720	−163.4	−37.5	15.2 %	58.7 %
Objectively defined (OBJ) threshold model				
5	−35.8	−2.0	32.6 %	45.7 %
10	12.1	−3.0	47.8 %	52.2 %
15	−35.8	−2.0	34.8 %	45.7 %
30	4.2	−3.0	50.0 %	54.3 %
60	12.0	−1.4	15.2 %	34.8 %
180	9.3	0.0	15.2 %	41.3 %
360	−27.2	−7.0	8.7 %	32.6 %
720	−41.1	−12.0	32.6 %	45.7 %

Negative values indicate threshold exceedance prior to the recorded time of debris flow, positive values indicate threshold exceedence after the recorded time of debris flow

poorly understood relative to other forms of landslide and debris-flow initiation. We recommend that future research focus upon improving our understanding (1) of the physical mechanisms that link short bursts of high intensity rainfall to debris-flow initiation, and (2) the relation between critical values of triggering rainfall intensities and physiographic characteristics, such as soil properties, surface hydrology, antecedent condition, and basin morphology.

Conclusions

This study described a method for assessing and improving the forecasting ability for debris flows in recently burned watersheds. We have improved upon the threshold delineation approach of Cannon et al. (2008) by (1) analyzing rainfall intensity data that characterizes the conditions that occur immediately prior to the timing of a debris-flow event, rather than peak rainstorm intensity, (2) defining thresholds using an objective approach that balances predictive successes with false alarms and failed alarms, and (3) improving the temporal correlation between the time at which rainfall intensity thresholds are exceeded and the timing of debris-flow events.

Our precise debris-flow timing database derived from 43 events, combined with rainfall intensity data collected within close proximity represents the most comprehensive single database of post-fire debris-flow initiation conditions to date. The literature review of Guzzetti et al. (2008) identified that only 133 of the 2626 documented debris flow events were able to constrain timing to ±12 h. Our study is able to constrain the exact rainfall conditions which contribute to the initiation of post-fire debris flows using rapidly deployable yet robust monitoring equipment. These data provide a means of comparing the traditional, peak-intensity approaches to ID threshold definition with the actual rainfall intensities occurring during debris flow initiation. We have identified that empirical thresholds based on peak rainstorm intensity are too high relative to the meteorological conditions prior to the

recorded time of debris flows. This finding does not diminish the value or need for empirical thresholds based on peak-storm intensity, because precise timing data for landslides and debris flows is typically unavailable or in limited quantity (Guzzetti et al. 2008). However, thresholds defined for the prediction of post-fire debris-flow events should be derived from precise timing data whenever possible, particularly when used for forecasting purposes, and post-fire (and other landslide and debris flow) monitoring efforts should emphasize collection of debris-flow and landslide timing data for future refinement of thresholds defined from peak rainstorm intensities. Our data, in combination with a quantitative and objective method provides a new and computationally simple means of defining the threshold conditions for the initiation of post-fire debris flows. Furthermore, this method allows us to quantitatively evaluate the predictive ability of existing thresholds during subsequent events, and modify rainfall thresholds throughout the winter season as necessary.

Given that rainfall ID thresholds constitute a significant decision-making factor for the southern California flash-flood and debris-flow early warning system (Restrepo et al. 2008), lowering threshold to the LL model and increasing the number of false alarms may be detrimental to the perceived effectiveness of the system (Baum and Godt 2010). Setting a rainfall threshold too low will also result in unnecessary and expensive evacuations and resource deployments, while thresholds that are too high could lead to a false sense of security and the possibility for injuries and fatalities. However, local, state and federal emergency management agencies, public works departments and weather forecasters are not likely to view false alarms as entirely negative, as these provide opportunities as learning experiences which allow them to modify and improve response strategies and contingency plans (Barnes et al. 2007). Nevertheless, over-warning the public may result in an increase in complacency towards debris-flow hazards. We feel that the objectively defined thresholds based upon the presented method offer the best balance between false and failed alarm rates within an early warning system. Other users may wish to incorporate a weighting coefficient in the calculation of a TS if they require unequal weights between false alarms and failed alarms.

Our results also highlight the importance of accurate forecasting of rainfall conditions for successful early warning of debris flows in recently burned watersheds. Since short-duration thresholds proved to be most effective for predicting debris-flow initiation, real-time monitoring of rainfall intensities within the burned watersheds provides little or no lead time for warning or evacuation. Improvement of rainfall intensity forecasting and monitoring of rainfall conditions upwind of the area of concern are of utmost importance for early warning of debris flows in recently burned watersheds. Recent improvements in radar technology, such as dual-polarization (Jorgensen et al. 2011), may improve forecasters' ability to resolve areas of high intensity rainfall within a storm and provide better guidance to emergency management teams and public dissemination of short lead-time warning.

Finally, the methods presented here related to the use of precise debris-flow timing data and utility functions to objectively define and evaluate critical rainfall intensity thresholds may be used to develop empirical thresholds for other types of landslides, or in other regions. Since our data and results are specific to the study basins in two burn areas in southern California, the objective thresholds defined in our analysis are likely only applicable in this particular region. However, the methods we have documented have practical applications not just in landslide and debris flow research, but also in the research of any system that exhibits nonlinear behavior characterized by critical thresholds.

Acknowledgements

This research was made possible with funding from the U.S. Geological Survey Landslide Hazards Program and Multi-Hazards Demonstration Project. The authors are grateful for field assistance from Joseph Gartner (USGS), Maiana Hanshaw (USGS), and Robert Leeper (USGS), rainfall data provided by Pete Wohlgemuth (U.S. Forest Service), and data processing by Matthew Hansen (University of Colorado–Denver). Jonathan Godt, Rex Baum, and two anonymous reviewers have provided comments and suggested revisions which have greatly improved this paper.

References

Aleotti P (2004) A warning system for rainfall-induced shallow failures. Eng Geol 73(3–4):247–265

Arattano M, Marchi L (2000) Video-derived velocity distribution along a debris flow surge. Phys Chem Earth, Part B: Hydrol Oceans Atmos 25(9):781–784

Arattano M, Marchi L (2005) Measurements of debris flow velocity through cross-correlation of instrumentation data. Nat Hazards Earth Syst Sci 5:137–142

Barnes LR, Gruntfest EC, Hayden MH, Schultz DM, Benight C (2007) False alarms and close calls: a conceptual model of warning accuracy. Weather Forecast 22(5):1140–1147

Baum RL, Godt JW (2010) Early warning of rainfall-induced shallow landslides and debris flows in the USA. Landslides 7(3):259–272

Baum RL, Godt JW, Savage WZ (2010) Estimating the timing and location of shallow rainfall-induced landslides using a model for transient, unsaturated infiltration. J Geophys Res 115(F3):F03013

Berti M, Genevois R, Simoni A, Tecca PR (1999) Field observations of a debris flow event in the Dolomites. Geomorphology 29(3–4):265–274

Bracchini M, Zannoni A (2003) Relations between rainfall and triggering of debris-flow: case study of Cancia (Dolomites, Northeastern Italy). Nat Hazard Earth Syst Sci 3(1):71–79

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(3):447–458

Caine N (1980) The rainfall intensity: duration control of shallow landslides and debris flows. Geogr Ann Ser A Phys Geogr 62(1–2):23–27

Cannon S, Boldt E, Laber J, Kean J, Staley D (2011) Rainfall intensity–duration thresholds for postfire debris-flow emergency-response planning. Nat Hazard 59(1):209–236

Cannon SH, Boldt EM, Kean JW, Laber J, Staley DM (2010) Relations between rainfall and postfire debris-flow and flood magnitudes for emergency-response planning, San Gabriel Mountains, southern California. U S Geol Surv Open File Rep 2010–1039:31pp.

Cannon SH, Ellen SD (1988) Rainfall activity that resulted in abundant debris-flow activity during the storm. In: Ellen SD, Wieczorek GF (eds) Landslides, floods and marine effects of the storm of January 3–5, 1982, in the San Francisco Bay Region, California. U.S. Geological Survey Profesional Paper 1434. pp 27–34

Cannon SH, Gartner JE, Parrett C, Parise M (2003) Wildfire-related debris-flow generation through episodic progressive sediment-bulking processes, western USA. In: Rickenmann D, Chen CL (eds) Debris-flow hazards mitigation — mechanics, prediction, and assessment, Proceedings of the Third International Conference on Debris-Flow Hazards Mitigation, Davos, Switzerland, 10–12 September 2003. Balkema, Rotterdam, pp 71–82

Cannon SH, Gartner JE, Rupert MG, Michael JA, Staley DM, Worstell BB (2009) Emergency assessment of postfire debris-flow hazards for the 2009 Station Fire, San Gabriel Mountains, southern California. U S Geological Survey Open File Report 2008–1370: 20pp

Cannon SH, Gartner JE, Wilson R, Bowers J, Laber J (2008) Storm rainfall conditions for floods and debris flows from recently burned areas in southwestern Colorado and southern California. Geomorphology 96(3–4):250–269

Cannon SH, Kirkham RM, Parise M (2001) Wildfire-related debris-flow initiation processes, Storm King Mountain, Colorado. Geomorphology 39:171–188

Cannon SH, Reneau SL (2001) Conditions for generation of fire-related debris flows, Capulin Canyon, New Mexico. Earth Surf Process Landforms 25:1103–1121

Carrara A, Crosta G, Frattini P (2008) Comparing models of debris-flow susceptibility in the alpine environment. Geomorphology 94(3–4):353–378
Cervi F, Berti M, Borgatti L, Ronchetti F, Manenti F, Corsini A (2010) Comparing predictive capability of statistical and deterministic methods for landslide susceptibility mapping: a case study in the northern Apennines (Reggio Emilia Province, Italy). Landslides 7(4):433–444
Chawner WD (1935) Alluvial fan flooding: the Montrose, California flood of 1934. Geogr Rev 25:235–263
Chien-Yuan C, Tien-Chien C, Fan-Chieh Y, Wen-Hui Y, Chun-Chieh T (2005) Rainfall duration and debris-flow initiated studies for real-time monitoring. Environ Geol 47 (5):715–724
Coe J, Kinner D, Godt J (2008) Initiation conditions for debris flows generated by runoff at Chalk Cliffs, central Colorado. Geomorphology 96(3–4):270–297
Crozier MJ (1999) Prediction of rainfall-triggered landslides: a test of the antecedent water status model. Earth Surf Process Landforms 24:825–833
Eaton EC (1935) Flood and erosion control problems and their solution. Proc Am Soc Civ Eng 62(8):1302–1362
Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27:861–874
Frattini P, Crosta G, Sosio R (2009) Approaches for defining thresholds and return periods for rainfall-triggered shallow landslides. Hydrol Processes 23(10):1444–1460
Godt JW, Baum RL, Chleborad AF (2006) Rainfall characteristics for shallow landsliding in Seattle, Washington, USA. Earth Surf Process Landforms 31(1):97–110
Godt JW, Baum RL, Savage WZ, Salciarini D, Schulz WH, Harp EL (2008) Transient deterministic shallow landslide modeling: Requirements for susceptibility and hazard assessments in a GIS framework. Eng Geol 102(3–4):214–226
Godt JW, McKenna JP (2008) Numerical modeling of rainfall thresholds for shallow landsliding in the Seattle, Washington area. In: Baum RL, Godt JW, Highland LM (eds) Landslides and engineering geology of the Seattle, Washington Area. Geol Soc Am Rev Eng Geol XX:121–135
Gregoretti C, Fontana GD (2007) Rainfall thresholds for the initiation of debris flow by channel-bed failure in the Dolomites. In: Chen C-I, Major JJ (eds) Debris-flow hazards mitigation; mechanics, prediction, and assessment. Proceedings of the 4th International Conference on Debris-Flow Hazards Mitigation, vol 4. pp 11–21
Gregoretti C, Fontana GD (2008) The triggering of debris flow due to channel-bed failure in some alpine headwater basins of the Dolomites: analyses of critical runoff. Hydrol Processes 22(13):2248–2263
Guzzetti F, Peruccacci S, Rossi M, Stark CP (2007) Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorology and Atmospheric Physics 98(3–4):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(1):3–17
Hong Y, Adler R, Huffman G (2006) Evaluation of the potential of NASA multi-satellite precipitation analysis in global landslide hazard assessment. Geophys Res Lett 33(22):L22402
Iiritano G, Versace P, Sirangelo B (1998) Real-time estimation of hazard for landslides triggered by rainfall. Environ Geol 35(2):175–183
Johnson AM, Rodine JR (1984) Debris flow. In: Brundsen D, Prior DB (eds) Slope instability. John Wiley and Sons, New York, pp 257–361
Jorgensen DP, Hanshaw MN, Schmidt KM, Laber JL, Staley DM, Kean JW, Restrepo PJ (2011) Value of a dual-polarized gap-filling radar in support of southern California post-fire debris-fow warnings. J Hydrometeorol 12(6):1581–1595
Kean JW, Staley DM, Cannon SH (2011) In situ measurements of post-fire debris flows in southern California: comparisons of the timing and magnitude of 24 debris-flow events with rainfall and soil moisture conditions. J Geophys Res 116(F4):F04019
Kean JW, Staley DM, Leeper RJ, Schmidt KM (2012) A low-cost method to measure the timing of post-fire flash floods and debris flows relative to rainfall. Water Resour Res 48(5):W05516
Larsen MC, Simon A (1993) A rainfall intensity–duration threshold for landslides in a humid-tropical environment, Puerto Rico. Geografiska Annaler Series A, Physical Geography 75(1–2):13–23
Martelloni G, Segoni S, Fanti R, Catani F (2012) Rainfall thresholds for the forecasting of landslide occurrence at regional scale. Landslides :1–11
McCoy SW, Kean JW, Coe JA, Staley DM, Wasklewicz TA, Tucker GE (2010) Evolution of a natural debris flow: in situ measurements of flow dynamics, video imagery, and terrestrial laser scanning. Geology 38(8):735–738
McPhee JA (1989) The control of nature. Farrar, Straus and Giroux, New York
Montesarchio V, Ridolfi E, Russo F, Napolitano F (2011) Rainfall threshold definition using an entropy decision approach and radar data. Nat Hazards Earth Syst Sci 11(7):2061–2074
Montgomery DR, Dietrich WE (1994) A physically based model for the topographic control on shallow landsliding. Water Resour Res 30(4):1153–1171
Morton DM (2004) Preliminary digital geological map of the 30′ × 60′ Santa Ana Quadrangle, southern California, version 2.0. US Geological Survey Open File Report 99–172
Okano K, Suwa H, Kanno T (2012) Characterization of debris flows by rainstorm condition at a torrent on the Mount Yakedake volcano, Japan. Geomorphology 136(1):88–94
Ponziani F, Pandolfo C, Stelluti M, Berni N, Brocca L, Moramarco T (2012) Assessment of rainfall thresholds and soil moisture modeling for operational hydrogeological risk prevention in the Umbria region (central Italy). Landslides, 1–9
Restrepo P, Jorgensen DP, Cannon SH, Costa J, Laber J, Major J, Martner B, Purpura J, Werner K (2008) Joint NOAA/NWS/USGS prototype debris flow warning system for recently burned areas in southern California. Bull Am Meteorol Soc 89(12):1845
Salciarini D, Tamagnini C, Conversini P, Rapinesi S (2011) Spatially distributed rainfall thresholds for the initiation of shallow landslides. Natural Hazards
Schaefer JT (1990) The critical success index as an indicator of warning skill. Weather Forecast 5:570
Schmidt KM, Hanshaw MN, Howle JF, Kean JW, Staley DM, Stock JD, Bawden GW (2011) Hydrologic conditions and terrestrial laser scanning of post-fire debris flows in the San Gabriel Mountains, CA, USA. In: Genevois R, Hamilton DL, Prestininzi A (eds) Proceedings of the Fifth International Conference on Debris Flow Hazards Mitigation/ Mechanics, Prediction, and Assessment, Padua, Italy, June 7–11, 2011. Italian Journal of Engineering Geology and Environment—Book: Casa Editrice Universita La Sapienza, Rome, Italy, pp 583–593
Schumm SA (1973) Geomorphic thresholds and complex response of drainage systems. In: Morrisawa M (ed) Fluvial geomorphology. State University of New York, Binghamton, NY, pp 299–310
Scott KM (1971) Origin and sedimentology of 1969 debris flows near Glendora, California. In: Geological Survey Research 1971, Chapter C, U.S. Geological Survey Professional Paper 750C. ppC242–C247
Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240(4857):1285–1293
Takahashi T (1981) Debris flow. Annu Rev Fluid Mech 13:57–77
Tang C, Rengers N, van Asch TWJ, Yang YH, Wang GF (2011) Triggering conditions and depositional characteristics of a disastrous debris flow event in Zhouqu city, Gansu Province, northwestern China. Nat Hazards Earth Syst Sci 11(11):2903–2912
Tecca P, Genevois R (2009) Field observations of the June 30, 2001 debris flow at Acquabona (Dolomites, Italy). Landslides 6(1):39–45
Wieczorek GF, Guzzetti F (2000) A review of rainfall thresholds for triggering landslides. In: Claps P, Siccardi F (eds) Mediterranean storms. Proceedings Plinius Conference '99, Maratea, 14–16 October 1999, vol GNDCI pub. n. 2010. Editoriale Bios, Cosenza, pp 407–414
Wilcoxon F (1945) Individual comparisons by ranking methods. Biom Bull 1(6):80–83
Yerkes RF, Cambell RH (2005) Preliminary geologic map of the Los Angeles 30′×60′ Quadrangle, southern California. U S Geological Survey Open File Report 2005–1019
Yong L, Jingjing L, Kaiheng H, Pengcheng S (2012) Probability distribution of measured debris-flow velocity in Jiangjia Gully, Yunnan Province, China. Nat Hazard 60(2):689–701

D. M. Staley (✉) · J. W. Kean · S. H. Cannon
U.S. Geological Survey,
Box 25046MS966 DFC, Denver, CO 80225, USA
e-mail: dstaley@usgs.gov

K. M. Schmidt
U.S. Geological Survey,
Menlo Park, CA, USA

J. L. Laber
National Oceanic and Atmospheric Administration,
National Weather Service,
Oxnard, CA, USA