Comparing threshold definition techniques for rainfall induced landslides: a national assessment using radar rainfall

EARTH SURFACE PROCESSES AND LANDFORMS
Earth Surf. Process. Landforms **43**, 553–560 (2018)

Published online 17 August 2017 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/esp.4202

Letters to ESEX

Comparing threshold definition techniques for rainfall-induced landslides: A national assessment using radar rainfall

Benjamin Postance,[1]* John Hillier,[1] Tom Dijkstra[2] and Neil Dixon[3]
[1] Department of Geography, Loughborough University, Loughborough, UK
[2] British Geological Survey, Keyworth, Nottingham, UK
[3] Civil and Building Engineering, Loughborough University, Loughborough, UK

Received 30 January 2017; Revised 1 June 2017; Accepted 10 July 2017

*Correspondence to: Benjamin Postance. E-mail: b.postance@lboro.ac.uk

ESPL
Earth Surface Processes and Landforms

ABSTRACT: Translational landslides and debris flows are often initiated during intense or prolonged rainfall. Empirical thresholds aim to classify the rain conditions that are commonly associated with landslide occurrence and therefore improve understating of these hazards and predictive ability. Objective techniques that are used to determine these thresholds are likely to be affected by the length of the rain record used, yet this is not routinely considered. Moreover, remotely sensed spatially continuous rainfall observations are under-exploited. This study compares and evaluates the effect of rain record length on two objective threshold selection techniques in a national assessment of Scotland using weather radar data. Thresholds selected by 'threat score' are sensitive to rain record length whereas, in a first application to landslides, 'optimal point' (OP) thresholds prove relatively consistent. OP thresholds increase landslide detection and may therefore be applicable in early-warning systems. Thresholds combining 1- and 12-day antecedence variables best distinguish landslide initiation conditions and indicate that Scottish landslides may be initiated by lower rain accumulation and intensities than previously thought. © 2017 The Authors. Earth Surface Processes and Landforms published by John Wiley & Sons Ltd.

KEYWORDS: landslides; rainfall; thresholds; early warning; debris flow

Introduction

Observational data suggest a positive relationship between rainfall and the occurrence of translational landslides and debris flows, hereafter 'landslides' (Iverson, 1997, 2000; De Vita *et al.,* 1998; Ballantyne, 2002). Rainfall-induced landslides pose a significant threat to life and often result in costly physical damage to property and the disruption of infrastructure, such as transport or power (Jaroszweski *et al.,* 2010; Meyer *et al.,* 2015; Postance *et al.,* 2017). To address these impacts and reduce risk requires an improved understanding of how landslide hazards are distributed in time and what rainfall conditions are relevant to landslide initiation. However, establishing these is non-trivial. Landscapes containing slopes that are susceptible to generating landslides can be regarded as being in equilibrium with the long-term environmental conditions, such as rainfall and land use, to which these have been exposed (Giannecchini *et al.,* 2016). It therefore follows that, when these conditions deviate from the long-term trend, a response in the landscape will be manifested. For landscapes where rainfall is a dominant factor, and other contributing factors such as land use remain relatively constant, it is reasonable to use rainfall analysis to quantify empirical thresholds for rain conditions at which hydrogeological responses are sufficient to disrupt the equilibrium and result in the triggering of a landslide (Caine, 1980; Van Asch *et al.,* 1999; Waltham and Dixon, 2000; Pennington *et al.,* 2014).

The earliest form of threshold focused on the characterization of the intensity and duration (ID) of rainfall events in which landslides occurred (Caine, 1980). This involves the graphical delineation of upper and lower limit ID thresholds to form an envelope of rainfall conditions within which landslides are likely to occur (Caine, 1980; Wieczorek and Glade, 2005; Guzzetti *et al.,* 2007; Cannon *et al.,* 2008). However, the visual selection of ID thresholds restricts ease of comparison and repeatability across study areas (Guzzetti *et al.,* 2007). Recently, other rain variables have been evaluated, such as antecedent rainfall accumulation, and more objective threshold selection techniques have been developed. These include selecting thresholds that capture different percentiles of the rain events

that initiated landslides (LRE) using a Bayesian inference model (Guzzetti *et al.,* 2007, 2008) or using thresholds drawn parallel to the log-transformed LRE line of best fit (Brunetti *et al.,* 2010; Meyer *et al.,* 2012; Gariano *et al.,* 2015). Other techniques explicitly consider the rain events that do not initiate landslides (NRE) as well. Thresholds with the greatest predictive accuracy are selected using receiver operating characteristic analysis (ROC) to maximize the number of correctly identified LRE and minimize the number of NRE and incorrect results (Jakob and Weatherly, 2003; Staley *et al.,* 2013; Segoni *et al.,* 2014b, 2015; Giannecchini *et al.,* 2016; Piciullo *et al.,* 2016). Alternatively, thresholds are selected that provide the greatest conditional probability of landslide occurrence and using analysis of daily rainfall time series rather than events (Chleborad, 2000; Chleborad *et al.,* 2008).

It is evident that for each of the above techniques their threshold values and probabilities may be influenced by uncertainties that restrict their application in landslide early-warning systems (Guzzetti *et al.,* 2008; Kirschbaum *et al.,* 2012). A principal source of this uncertainty is in the proportion of LRE and NRE used in the analysis. This proportion is determined by two subjective factors.

The first factor relates to the criteria used to define rain events. A common standard or set of criteria is lacking and there is little consistency between approaches in the literature (Melillo *et al.,* 2015). Several studies define LRE as the time period prior to landslide occurrence, ranging from 14 to 150 days, and use an equivalent number of NRE that are randomly selected or are of a similar accumulation magnitude (Jakob and Weatherly, 2003; Winter *et al.,* 2010; Meyer *et al.,* 2012; Søren *et al.,* 2014). Others use minimum dry period criteria to separate individual rain events. The criteria range from 2 to 120 h and are applied to rain records for periods of landslide activity lasting 16 months to several years (Saito *et al.,* 2010; Staley *et al.,* 2013) or using all available rain records up to 70 years in length (Brunetti *et al.,* 2010; Berti *et al.,* 2012; Marra *et al.,* 2014; Melillo *et al.,* 2015; Abancó *et al.,* 2016; Piciullo *et al.,* 2016). As a consequence, it is difficult to compare thresholds based on different event criteria (Melillo *et al.,* 2015). In addition, it is uncertain how applicable each criterion is in regions with different hydrological regimes to those where they were developed. Thresholds based on analysis of daily time series data avert this issue by including all available rain information at the expense of greater numbers of NRE – days without landslides (Chleborad, 2000; Chleborad *et al.,* 2008).

The second factor is related to the type and length of the rain record used. Thresholds are typically derived using rain gauge networks and a single gauge record (and hence set of LRE and NRE) that is associated with all of the landslides that occur within a region or to sub-domains based on the spatial distribution of gauges or landslide mechanisms (Chleborad, 2000; Guzzetti *et al.,* 2007; Chleborad *et al.,* 2008; Brunetti *et al.,* 2010; Segoni *et al.,* 2014a). Increasingly, thresholds are derived using remotely sensed data, including satellite (Hong *et al.,* 2006; Kirschbaum *et al.,* 2012) and high-resolution weather radar (Winter *et al.,* 2010; Marra *et al.,* 2014; Abancó *et al.,* 2016). These spatially continuous observations provide a site-specific rain record (SSR) at each landslide location. This is beneficial in the study of landslide initiation thresholds: (i) due to high spatial variability of rainfall, particularly in mountainous terrain (Sidle and Ochiai, 2013); and (ii) because they provide self-consistency between the measurements used to develop thresholds and those suited to early warning, such as national coverage and 36 h forecasts (Met Office, 2003; Borga *et al.,* 2014). Nonetheless, the use and length of SSR can lead to a significant increase in the number of days without landslides (i.e. NRE) and so influence threshold selection, including the sensitivity of threshold values and occurrence probabilities. These effects are yet to be systematically evaluated for different threshold selection techniques.

In this investigation thresholds are derived using radar SSR time series, as applied to gauge records and sub-domains by Chleborad *et al.* (2008). ROC analysis is used to select two types of threshold. First, 'threat score' (TS) is used to select the threshold that offers the greatest level of predictive accuracy and minimizes false results (Staley *et al.,* 2013). Second, the 'optimal point' (OP) threshold is selected to provide the highest rate of landslide detection for the lowest number of false results. OP is yet to be applied to landslides, but is common in other disciplines for the detection of rare phenomena, including disease and signal processing (Schisterman *et al.,* 2005; Perkins and Schisterman, 2006; Rota and Antolini, 2014). For each threshold, the effect of varying the SSR record length is systematically evaluated using 10 different record lengths representative of those implemented in other studies. A national landslide assessment of Scotland is used as a case study.

Case Study and Data

The Scottish mainland (70 100 $km^{2)}$ is characterized by mountains, upland moors and incised valleys that feature many steep slopes. At high elevations (≤1346 m above sea level) the bedrock is exposed whereas slopes are overlain by unconsolidated sandy regolith and postglacial deposits <5 m in depth (Ballantyne, 1986, 2002; Trewin, 2002). In such shallow, coarse-grained materials there is often a rapid response to rainfall, whereas finer materials have lower permeabilities or have deeper slip surfaces, which may see delays between 'triggering' rain conditions and landslide occurrence of several days or more (Iverson *et al.,* 1997; Stoffel *et al.,* 2014). Although there is evidence of large, complex landslides and rock slope failures such as at Trotternish, Skye (Ballantyne *et al.,* 2014), the impact to society is dominated by shallow slides and debris flow impacting the highway network during sequences of autumn and winter storms, including in 2004, 2007, 2009, 2012 and 2015 (Winter *et al.,* 2010; Postance *et al.,* 2017). The annual number of reported landslides in 'landslide seasons' for the period 2003–2016 varies from 4 to 5 to 43 events in 2012 (Foster *et al.,* 2008; Pennington and Harrison, 2013; Dijkstra *et al.,* 2016), yet many small failures are often unrecorded (Milne *et al.,* 2009).

This study uses a database of 75 landslides that occurred in the period 2004–2016 recorded in the British Geological Survey (BGS) National Landslide Database (NLDB) (Foster *et al.,* 2008). The NLDB includes the coordinates of landslide source locations and the date and time of failure (Figure 1); temporal accuracy is <12 h (20%), 24 h (57%), 48 h (7%) and unknown (16%). The data quality is typical of most landslide inventories, with bias towards larger events and those with impact to society (Malamud *et al.,* 2004; Guzzetti *et al.,* 2008). SSR data are obtained from a network of 15 C-band Doppler radars (Met Office, 2003), providing precipitation estimates at 5 km^2 resolution for 15 min intervals over the entire UK and Western Europe for the period March 2004 to present. The radar is calibrated to account for, among others, echoing, beam blockage and attenuation as described in Harrison *et al.* (2000, 2009). At ≥1 h aggregations the radar data adequately reproduce gauged precipitation observations, however, the accuracy for short-duration and low-intensity precipitation is uncertain as most gauges require ≥0.2 mm per record (Harrison *et al.,* 2000; Villarini and Krajewski, 2008; Villarini *et al.,* 2008). The radar data are widely applied for research on UK flooding (Schellart *et al.,* 2012; Parkes *et al.,* 2013), climatology (Fairman *et al.,* 2015) and water management (Harrison *et al.,* 2009).

Thresholds are developed using normalized rain variables, as thresholds that apply in one region do not apply to regions with different climatic regimes and weather variability (Jakob and

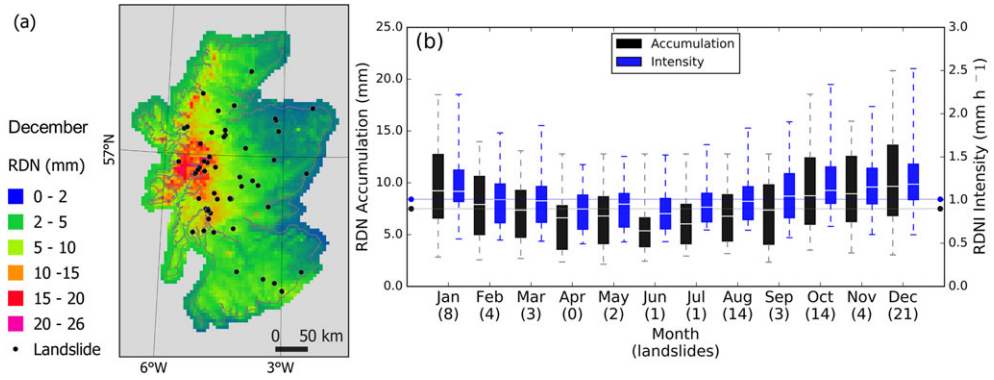

Figure 1. (a) Map of landslide locations (black dots) and December RDN values in Scotland (2004–2016). (b) A boxplot summarizing the monthly RDN (black) and RDNI (blue) values for 75 landslide locations. The boxplots are median cantered and range from first to third quartiles with whiskers to minimum and maximum values. The *x*-axis labels indicate the number of landslides recorded in each month. The points and horizontal lines show the average annual RDN (black) and RDNI (blue) for comparison. [Colour figure can be viewed at wileyonlinelibrary.com]

Weatherly, 2003). For each location and pixel, the rainy-day normal (RDN mm = *V*/*D*) is the average of total rain accumulation (*V*) and days with >0.1 mm rain (*D*) in a year (Wilson and Jayko, 1997). RDN indicates typical rainfall accumulation by reducing the effect of no-rain days (Brunetti *et al.*, 2010; Meyer *et al.*, 2012; Gariano *et al.*, 2015). Illustrated in Figure 1a and Figure 1b, monthly RDN values are used to account for seasonal rainfall patterns. The rain variables used are: non-normalized, total rain accumulation (*V*, mm); RDN normalized rain accumulation (NV = *V*/RDN); hourly maximum rain intensity (I_{max}, mm h^{-1}); and rain duration (RD, h). RD is a new measure for the hours of rainfall above the monthly rainy-day normal intensity (RDNI = *I*/*t*). Similar to RDN, RDNI is the average of the rainfall intensities (*I*) recorded for all time periods (*t*) with >0.01 mm h^{-1} rain (Figure 1b). Each variable is calculated for antecedent periods of 1, 2, 3, 6, 12, 18, 24, 30, 50 and 60 days (e.g. 12 days = NV_{12}).

Method

The determination of thresholds can be approached as a binary classification problem, where rainfall conditions coinciding with landslides are separated from rain conditions that do not coincide with landslides. As illustrated in Figure 2, binary classification yields four mutually exclusive contingencies and ROC analysis is used to assess a classifier's performance, including for landslide thresholds (Wilks, 2006; Gariano *et al.*, 2015; Abancó *et al.*, 2016; Giannecchini *et al.*, 2016; Piciullo *et al.*, 2016). For each rain variable, the contingencies are calculated and an ROC curve is formed by plotting the false and true positive rate, where each point on the curve represents a different threshold value. The area under the curve (AUC) indicates the overall performance of a rain variable and a diagonal curve indicates randomness (AUC = 0.5) and therefore indiscriminate of landslides (Wilks, 2006). These measures are calculated for each individual rain variable and antecedent period, and for pairs of variables such as rain intensity and duration. Independent pairs are selected using a 50-fold cross-validated Spearman's rank correlation test (*R*) and the AUC measures.

Two types of threshold are selected. First, Staley *et al.* (2013) demonstrate that landslide thresholds can be selected by maximizing TS. As illustrated in Figure 2b, TS measures threshold accuracy when correct negatives are removed and ranges from 0 to 1 with each incorrect classification reducing TS (Schaefer, 1990). In ROC space (Figure 2a) the OP represents a hypothetical threshold that classifies all outcomes correctly. Second, therefore, thresholds are selected with minimum distance to OP to maximize landslide detection while minimizing errors. In other fields, OP thresholds are applied to the detection of rare phenomena and in situations where missed alarms (false negatives) may result in fatalities or unacceptable losses (Schisterman *et al.*, 2005; Rota and Antolini, 2014). OP

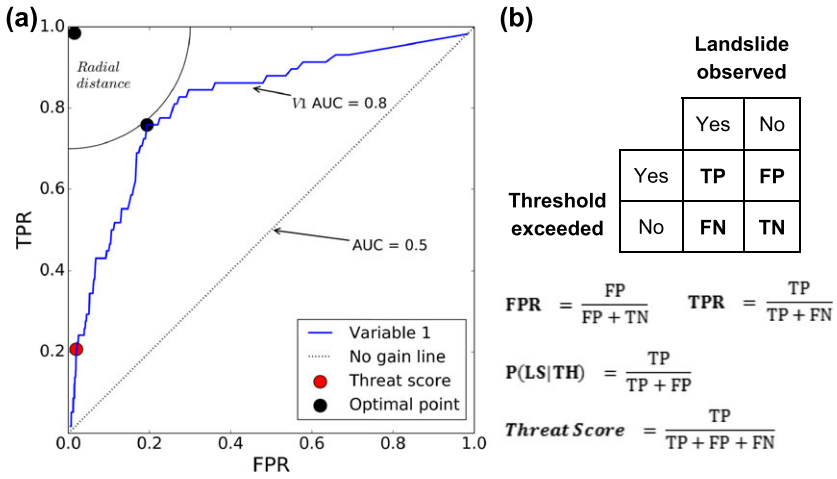

Figure 2. (a) Example ROC curve for a rain variable (V1, blue line). AUC indicates the predictive performance of a variable, and the no gain line (AUC = 0.5) indicates a variable with random performance. The two ROC threshold selection metrics are shown: (i) optimal point (black points); and (ii) threat score. (b) Is a contingency table and formulae for true positive rate (TPR), false positive rate (FPR), conditional probabilities of landslide occurrence *P*(LS|TH) and threat score. [Colour figure can be viewed at wileyonlinelibrary.com]

thresholds may therefore be applicable to landslide early warning but are yet to be applied. Threshold values and the conditional probability of landslide occurrence, *P*(LS|TH), is reported. The statistical significance of threshold probabilities is assessed using the binomial distribution (i.e. coin toss experiment) and only the significant thresholds are reported ($\alpha = 0.05$).

To evaluate the influence of rain record length, the contingency scores, ROC metrics, threshold values and their probabilities are calculated and compared using 10 different SSR record lengths that reflect the approaches of previous studies. These include daily SSR time series of 10, 20, 30, 45, 90, 180, 365 and 730 days before each landslide event, all days before ('Prior') and all ('All') which include all available data either side of landslide occurrence (March 2004 to August 2016). The 'Prior' record is influenced by the date of landslide occurrence.

Results

Table I displays the ROC measures, OP and TS thresholds averaged over each of the different rain record lengths for a selection of high-performing single rain variables. Comprehensive results are supplied as supporting information in the supplementary material. The values in parentheses show the change in threshold values and probabilities between the shortest and longest rain records ('10-day'–'All'). The coefficient of variance (CV) is a measure for the relative variance of the threshold values across the different rain record lengths.

Table I also displays results for pairs of rain variables. There are 595 possible pairs which have a mean Spearman correlation of $R = 0.52$ (range 0.17–0.99). The criteria $R < 0.40$ and mean AUC > 0.85 are applied to select pairs with relatively weak correlation and high classification performance. Twelve pairs were identified for different combinations of rain duration ($RD_{1,2-12}$), normalized rain accumulation ($NV_{1,2,12-18}$) and maximum rain intensity ($I_{max\ 1-2}$). Figures 3a and Figure 3b illustrate the TP and OP thresholds and their probabilities for the combination of RD_1 and NV_{12} and the effect of using the different rain record lengths.

Discussion

The AUC and threshold probabilities are highest for each rain variable at 1- and 2-day antecedence, with decreasing values at longer antecedence periods. On the day of landslide occurrence, the OP and TS thresholds for normalized rain accumulation (NV_1) range from 1.9 to 6.3 times the RDN, respectively. Adjusting by the monthly RDN values the thresholds translate to critical accumulations of approximately 11.0–36.5 mm in June, the driest month, and 19.6–65.0 mm in the wettest month December. This demonstrates how monthly RDN values are used to account for seasonal fluctuations of rainfall in Scotland. Moreover RDN, and thus thresholds, also vary spatially (Figure 1a). For NV_2, thresholds are 3.7–9.6 times RDN and, together with NV_1, quantify the significant fluxes in water supply prior to landslide occurrence. The 1- and 2-day antecedence RD thresholds indicate that critical durations of above-average rainfall are in the range of 3.5–30 h. The new thresholds are consistent with physical descriptions in Scotland, noting that most landslides occur on slopes featuring a shallow cover (<3 m) of relatively coarse granular matrix over impermeable substrata that

Table I. Rainfall thresholds for the possible initiation of landslides in Scotland

			Optimal point			Threat score		
Variable	*R*	AUC	Threshold	CV (%)	*p*(LS|TH)	Threshold	CV (%)	*p*(LS|TH)
Individual variables								
$I_{max\ 1}$		0.83	4.7 (5.6–4.4)	10	0.05 ($0.18–9.6 \times 10^{-4}$)	12.4 (5.6–14.1)	23	0.11 ($0.18–5.2 \times 10^{-3}$)
$I_{max\ 2}$		0.80	5.5 (6.6–4.8)	15	0.04 ($0.16–7.3 \times 10^{-4}$)	13 (8.7–14.2)	14	0.09 ($0.19–3.0 \times 10^{-3}$)
RD_1		0.86	4.0 (5.5–3.5)	20	0.06 ($0.22–1.0 \times 10^{-3}$)	14.4 (8.3–20.3)	37	0.21 ($0.33–3.3 \times 10^{-2}$)
RD_2		0.85	7.9 (9.4–6.4)	11	0.06 ($0.20–9.1 \times 10^{-4}$)	22.6 (16.2–30.7)	31	0.18 ($0.35–2.1 \times 10^{-2}$)
NV_1		0.88	1.9 (2.6–1.9)	13	0.07 ($0.26–1.3 \times 10^{-3}$)	6.3 (3.6–9.1)	27	0.22 ($0.37–1.7 \times 10^{-2}$)
NV_2		0.87	3.7 (4.4–3.5)	07	0.07 ($0.26–1.3 \times 10^{-3}$)	9.6 (7.0–20.7)	44	0.22 ($0.43–3.2 \times 10^{-2}$)
NV_{12}		0.81	15.0 (19.7–13.7)	12	0.05 ($0.20–8.7 \times 10^{-4}$)	28.4 (22.8–46.8)	34	0.10 ($0.34–1.8 \times 10^{-2}$)
Pairs of variables								
* $I_{max\ 1}$ + RD_{12}	0.39	0.89	4.7_32.1 *(5.0–6.0) + (39.0–21.6)	28 28	0.08 ($0.26–1.7 \times 10^{-3}$)	12.0_49.2 (13.0–13.0) (28.6–96.4)	29 62	0.32 ($0.6–8.6 \times 10^{-2}$)
$I_{max\ 1}$ NV_{12}	0.36	0.89	2.4_19.0 (4.0–2.0) (22.6–13.0)	35 17	0.11 ($0.38–2.1 \times 10^{-3}$)	11.9_19.5 (2.0–13.0) (25.0–25.0)	29 25	0.30 ($0.45–3.2 \times 10^{-2}$)
RD_1 NV_{12}	0.39	0.91	3.5_17.4 (3.5–3.5) (22.0–15.0)	0 15	0.11 ($0.38–2.3 \times 10^{-3}$)	12.1_18.6 (3.5–20.3) (24.0–18.0)	28 43	0.41 ($0.75–5.5 \times 10^{-2}$)

Variables: rain variable used and antecedence period. *R*: cross-validated Spearman's rank-order correlation coefficient. Threshold: the mean threshold value across 10 different rain records; pairs are separated using '_' and parentheses show the threshold values for (10-day–'All') rain records. CV: threshold coefficient of variance across 10 different rain records. *p*(LS|TH): conditional probability of landslide occurrence for threshold exceedance; parentheses show the probabilities for ('10-day'–'All') rain records. For pairs of variables, (*) and (+) demonstrate the order of results in cells.

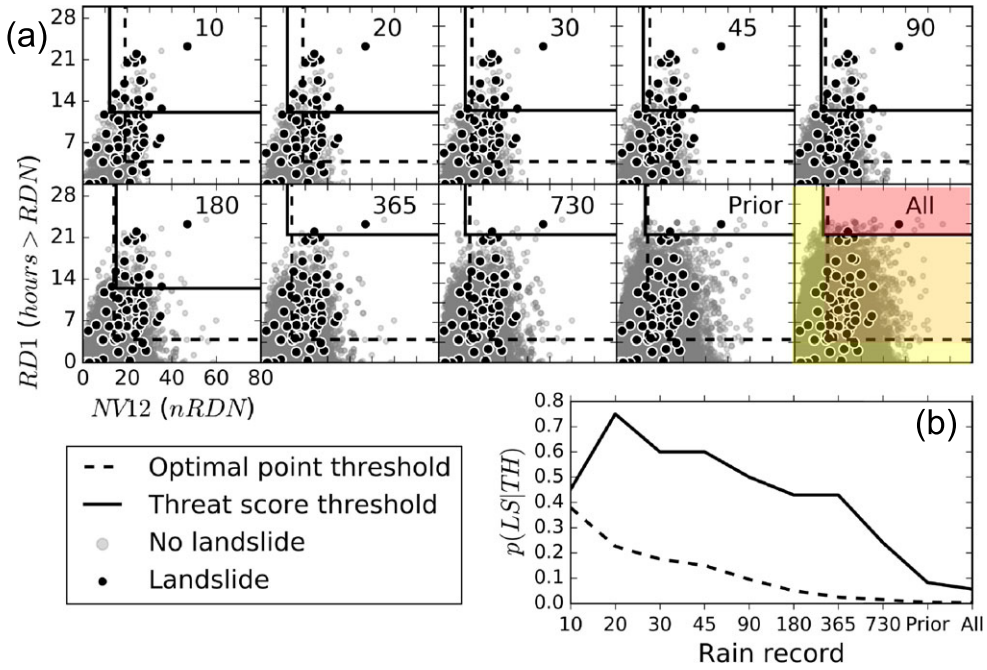

Figure 3. (a) Panel plot showing the OP (dashed line) and TS (solid line) thresholds for the combination of RD_1 and NV_{12} and for 10 different rain records. Black dots are the days with landslide occurrence and grey dots are days with no landslides. A hypothetical traffic light warning system (uncertain / low = yellow, moderate = amber, high = red) is shown for landslide thresholds on the 'All' rain record panel. (b) Plot of the conditional probability of landslide occurrence given OP or TS threshold exceedance for each rain record. [Colour figure can be viewed at wileyonlinelibrary.com]

are highly susceptible due to rapid water infiltration and saturation (Ballantyne, 1986, 2002; Milne *et al.,* 2009).

For I_{max}, similar threshold values of 4.7–5.5 mm h^{-1} (OP) and 12.4–13.0 mm h^{-1} (TS) are found for 1- and 2-day antecedence periods, respectively. These thresholds have lower performance measures relative to NV and RD, which may indicate that there are lags in water supply to reach critical parts of the slope – for instance, due to groundwater flow (Iverson, 1997, 2000) or increased surface runoff during high-intensity rain (Lu and Godt, 2013). Alternatively, in other regions initiation thresholds are found for intensities far below the peak intensity of rainstorms; however, these are for landslides initiated under post-fire conditions (Staley *et al.,* 2013). The relatively poor performance of I_{max} is more likely attributed to: (i) data artefacts, as 23% of the landslide records have >48 h temporal resolution; and (ii) spatiotemporal rainfall variations as I_{max} is a non-normalized rain variable.

A similar range of 1- and 2-day antecedence NV, RD and I_{max} thresholds are obtained in the pairs combining 12-day antecedent rainfall accumulation (NV_{12}) and produce improved performance measures (Table I). More robust thresholds are obtained as the antecedent accumulation results in the increase of slope material pore water pressure and a reduction of effective stress, so that shear strength is reduced and a landslide occurs (Iverson *et al.,* 1997; Wieczorek and Glade, 2005; Napolitano *et al.,* 2016). As shown in other empirical thresholds obtained in Scotland (Winter *et al.,* 2010), the results highlight the role of the 12-day antecedent hydrological condition of soils as a precursory factor controlling the rain thresholds that trigger landslides. Lower performance measures are obtained for combinations with longer antecedent accumulation periods ($\geq NV_{18}$) and this can be attributed to the geomorphological and geotechnical properties of soils, such as steep slope angles, high permeability and high granularity that contribute to rapid water drainage (Ballantyne, 1986; Wieczorek, 1996; Nettleton *et al.,* 2005). This also supports the observation that landslides occur more frequently during the early autumn and winter seasons, when sequences or 'clustering' of storm events is prevalent (Mailier *et al.,* 2006; Kendon and McCarthy, 2015; McCarthy *et al.,* 2016).

Combinations were also found with the 12-day rain duration variable (RD_{12}). However, RD is ambiguous for long antecedence periods as this may include the duration of a single rain event or the cumulative duration of multiple sub-events that resulted in landslides. Subsequently, for long antecedence periods it is more appropriate to distinguish the durations of individual rain events and the separating dry periods (Melillo *et al.,* 2015). This limitation restricts the suitability of the new RD variable to triggering events within the 1- to 2-day antecedence periods.

A common objective is to establish lower and upper limit thresholds that capture the rain conditions above and below which all landslides are known to have occurred (Guzzetti *et al.,* 2008). However, these are often derived irrespective of the non-landslide rain events (Caine, 1980; Guzzetti *et al.,* 2007; Brunetti *et al.,* 2010) or using approaches to maximize threshold accuracy, comparable to upper limit thresholds, at the expense of high landslide detection, such as TS (Staley *et al.,* 2013). As illustrated in Figure 3 and in Table I, the TS thresholds are relatively sensitive to rain record length and have a mean CV of 28% across all of the rain variables and pairs examined. Conversely, the mean CV for OP thresholds is 11% and is relatively consistent with respect to changes in the rain record length. This study also demonstrates landslide thresholds selected using SSR time series that are not influenced by the selection of suitable rain event criteria. However, the period of available radar observations is limited to 13 years, whereas gauge data are often available for longer periods of several decades. Nonetheless, changes to rain record length are shown to influence OP and TS thresholds differently and this therefore constitutes an important consideration for future investigations.

OP thresholds optimize landslide detection, limiting the number of false or missed alarms, and are therefore suggested as a suitable method to determine lower-limit initiation thresholds and which may be beneficial for landslide warning (Reid, 2006; Wachinger *et al.,* 2013; UNISDR, 2015). For example, Figure 3a illustrates a hypothetical three-tier (yellow–amber–red) early-warning system based on the 'All' derived thresholds for the RD_1 and NV_{12} pair. OP threshold exceedance is used for a moderate-possibility 'amber' warning to indicate the rain conditions in which most

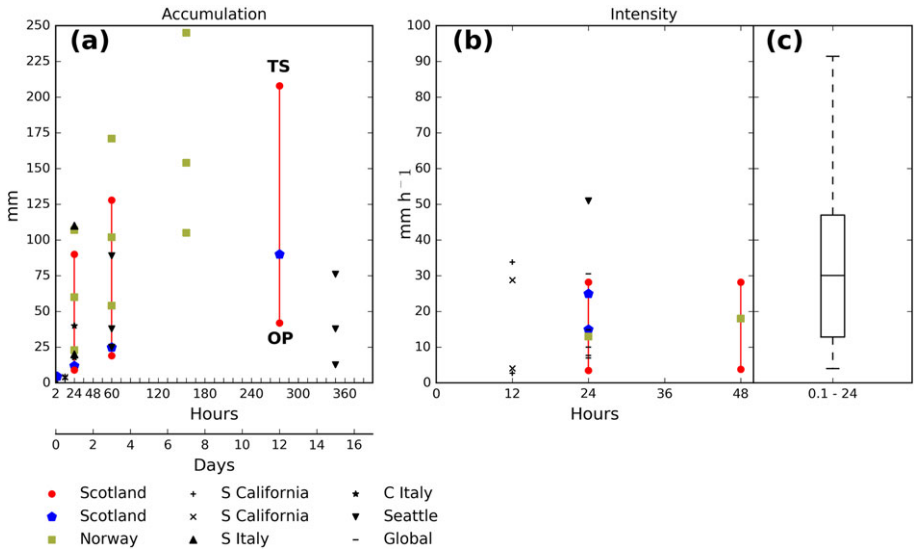

Figure 4. Comparison of new thresholds for Scotland (red dots and lines) and thresholds published in other temperate climates. (a) Rain accumulation thresholds from 0 to 17 days, and (b) rain intensity thresholds 0–48 h. Stacked symbols indicate studies reporting lower, middle and upper thresholds. Blue pentagons: Scotland (Winter *et al.,* 2010); green squares: Norway (Meyer *et al.,* 2012); cross: Southern California (Cannon *et al.,* 2008); X: Southern California (Staley *et al.,* 2013); triangle: southern Italy (Gariano *et al.,* 2015); star: central Italy (Peruccacci *et al.,* 2012); inverted triangle: Seattle (Chleborad *et al.,* 2008); dash: global (Guzzetti *et al.,* 2007). (c) Box whisker plot to show the range of 0.1–24 h maximum intensity thresholds for 25 different territories, including, Italy, Switzerland, Austria, Spain, North and South America, Indonesia, South East Asia and Japan (Guzzetti *et al.,* 2007). The box is median cantered and extends 1st to 3rd quartiles and whiskers 10% to 90%. Three thresholds are not shown at ~120 (2) and 180 (1) mm h^{-1}. [Colour figure can be viewed at wileyonlinelibrary.com]

landslides occur (64%). For completeness, a tentative yellow warning may be given below the OP threshold for uncertain or low-likelihood conditions while a red, high landslide likelihood warning is given for conditions that exceed the TS threshold. However, TS thresholds only capture a large proportion of landslides at ≤180 rain records and are more sensitive to rainfall record length (Figure 3a). The CV measure may be used to indicate the level of uncertainty for each threshold.

This study evaluates the effect of varying the rainfall record length, the primary data input, on two different threshold selection techniques. However, other investigations demonstrate that improved threshold performance may be achieved by considering additional sources of water on slopes, such as by snow melt or surface flow (Meyer *et al.,* 2012; Martelloni *et al.,* 2013; Søren *et al.,* 2014), and by using more sophisticated variables of effective precipitation due to evaporation and evapotranspiration (Dixon and Brook, 2007; Dijkstra and Dixon, 2010). In addition, further enhancements may be achieved by developing independent thresholds for sub-domains that are based on the spatial distribution of different slope lithologies that are associated with landslide occurrence (Segoni *et al.,* 2014a). Here, national scale thresholds are developed due to the relatively limited number of landslides recorded within the study area and as these records are comprised of similar landslide mechanisms and materials.

Figure 4 compares the new NV and I_{max} landslide thresholds with those published for other temperate regions and includes thresholds developed using: different threshold selection techniques; analysis of rainfall events for gauge and radar data; absolute and normalized rain variables; and output for lower, middle and upper limit thresholds. Several of the intensity thresholds may be inaccurate (±1–2 mm h^{-1}) due to the nature in which the thresholds are reported, for instance using mean or maximum values in graphics. The most relevant thresholds to this study are those for the analysis of 16 landslide rain events in Scotland (Winter *et al.,* 2010) and 502 events in Norway (Meyer *et al.,* 2012). For Scotland, the rain accumulation thresholds for 24 and 60 h (blue dots) correspond well with the OP thresholds of this study (red dots). Winter *et al.* (2010) also note the importance of 12-day antecedent accumulation; however, the new thresholds indicate lower 12-day accumulations and a wider range of initiation intensities. These differences are likely attributed to this study's use of more landslide records, thus greater variety of initiation conditions, and accumulation values normalized using monthly RDN values. For Norway, rain intensities are within the range of the OP and TS thresholds found for Scotland. The Norwegian rain accumulation thresholds are higher but these also include water supplied by melting ice and snow. Thresholds from other regions and countries worldwide have different, higher values, but this is likely attributed to different physical constraints and methodology of analysis. For instance, the $I_{max\ 1}$ OP thresholds are lower than much of the rest, but when looking at the range of intensities (OP–TS) these seem to correspond well with those reported for similar landslide mechanisms and climates (Figure 4c). This review indicates that it is important to regionally constrain comparisons.

Conclusions

This study examined two methods to determine objective landslide thresholds in an assessment of shallow translational slides and debris flows in Scotland. Thresholds were produced using information for 75 recorded events and radar precipitation data. The results show that thresholds selected using threat score are more sensitive to rainfall record length, whereas the optimal point thresholds provide rational lower limit values and these are least affected by the length of the rain record used – a factor not considered in past studies. Thresholds normalized by monthly rainy-day normal provide the most robust landslide thresholds, including the combination of rain duration on the day of failure ($RD_1 > 3.5$ h) and 12-day normalized accumulation ($NV_{12} > 17.4$), while maximum rain intensity thresholds perform relatively poorly.

Acknowledgements—The BGS NLDB is available online at http://www.bgs.ac.uk/geoindex/. The authors wish to thank the BGS Landslides team for assistance with the NLDB. NIMROD data are available online at http://www.ceda.ac.uk/ and commercially via the Met Office. This work was supported by the Natural Environment Research Council (grant number 1401793).

References

Abancó C, Hürlimann M, Moya J, Berenguer M. 2016. Critical rainfall conditions for the initiation of torrential flows: results from the Rebaixader catchment (Central Pyrenees). *Journal of Hydrology* **541**: 218–229.

Ballantyne CK. 1986. Landslides and slope failures in Scotland: a review. *Scottish Geographical Magazine* **102**: 134–150.

Ballantyne CK. 2002. Debris flow activity in the Scottish Highlands: temporal trends and wider implications for dating. *Studia Geomorphologica Carpatho-Balcanica* **36**: 7–27.

Ballantyne CK, Sandeman GF, Stone JO, Wilson P. 2014. Rock-slope failure following Late Pleistocene deglaciation on tectonically stable mountainous terrain. *Quaternary Science Reviews* **86**: 144–157.

Berti M, Martina MLV, Franceschini S, Pignone S, Simoni A, Pizziolo M. 2012. Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. *Journal of Geophysical Research - Earth Surface* **117**(4). https://doi.org/10.1029/2012JF002367.

Borga M, Stoffel M, Marchi L, Marra F, Jakob M. 2014. Hydrogeomorphic response to extreme rainfall in headwater systems: flash floods and debris flows. *Journal of Hydrology* **518**(PB): 194–205.

Brunetti MT, Peruccacci S, Rossi M, Luciani S, Valigi D, Guzzetti F. 2010. Rainfall thresholds for the possible occurrence of landslides in Italy. *Natural Hazards and Earth System Sciences* **10**: 447–458.

Caine N. 1980. The rainfall intensity: duration control of shallow landslides and debris flows. *Geografiska Annaler. Series A, Physical Geography* **62**: 23–27.

Cannon SH, Gartner JE, Wilson RC, Bowers JC, Laber JL. 2008. Storm rainfall conditions for floods and debris flows from recently burned areas in southwestern Colorado and southern California. *Geomorphology* **96**: 250–269.

Chleborad AF. 2000. Preliminary method for anticipating the occurrence of precipitation-induced landslides in Seattle, Washington. Open File Report 2000–469, US Geological Survey, Reston, VA.

Chleborad AF, Baum RL, Godt JW, Powers PS. 2008. A prototype system for forecasting landslides in the Seattle, Washington, area. *Reviews in Engineering Geology* **20**: 103–120.

De Vita P, Reichenbach P, Bathurst JC, Borga M, Crozier GM, Glade T, Guzzetti F, Hansen A, Wasowski J. 1998. Rainfall-triggered landslides: a reference list. *Environmental Geology* **35**: 219–233.

Dijkstra TA, Dixon N. 2010. Climate change and slope stability in the UK: challenges and approaches. *Quarterly Journal of Engineering Geology and Hydrogeology* **43**: 371–385.

Dijkstra TA, Reeves H, Freeborough K, Dashwood C, Pennington C, Jordan H, Hobbs P, Richardson J, Banks V, Cole S, Wells S, Moore R. 2016. Landslides in Scotland: case studies, threshold models and early warning. In *British Geological Survey Commercial Report, CR/16/206*. UK: Keyworth.

Dixon N, Brook E. 2007. Impact of predicted climate change on landslide reactivation: case study of Mam Tor, UK. *Landslides* **4**: 137–147.

Fairman JG, Schultz DM, Kirshbaum DJ, Gray SL, Barrett AI. 2015. A radar-based rainfall climatology of Great Britain and Ireland. *Weather* **70**: 153–158.

Foster C, Gibson A, Wildman G. 2008. The new National Landslide Database and Landslide Hazard Assessment of Great Britain. In First World Landslide Forum, Tokyo, Japan; 203–206.

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.

Giannecchini R, Galanti Y, D'Amato Avanzi G, Barsanti M. 2016. Probabilistic rainfall thresholds for triggering debris flows in a human-modified landscape. *Geomorphology* **257**: 94–107.

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**: 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.

Harrison D, Driscoll SJ, Kitchen M. 2000. Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorological Applications* **6**: 135–144.

Harrison DL, Kitchen M, Scovell RW. 2009. High-resolution precipitation estimates for hydrological uses. *Proceedings of the ICE – Water Management* **162**: 125–135.

Hong Y, Alder R, Huffman G. 2006. Evaluation of the potential of NASA multi-satellite precipitation analysis in global landslide hazard assessment. *Geophysical Research Letters* **33** L22402: doi:https://doi.org/10.1029/2006GL028010.

Iverson RM. 1997. The physics of debris flows. *Reviews of Geophysics* **35**: 245–296.

Iverson RM. 2000. Landslide triggering by rain infiltration. *Water Resources Research* **36**: 1897–1910.

Iverson RM, Reid ME, LaHusen RG. 1997. Debris-flow mobilization from landslides 1. *Annual Review of Earth and Planetary Sciences* **25**: 85–138.

Jakob M, Weatherly H. 2003. A hydroclimatic threshold for landslide initiation on the North Shore Mountains of Vancouver, British Columbia. *Geomorphology* **54**: 137–156.

Jaroszweski D, Chapman L, Petts J. 2010. Assessing the potential impact of climate change on transportation: the need for an interdisciplinary approach. *Journal of Transport Geography* **18**: 331–335.

Kendon M, McCarthy M. 2015. The UK's wet and stormy winter of 2013/2014. *Weather* **70**: 40–47.

Kirschbaum DB, Adler R, Hong Y, Kumar S, Peters-Lidard C, Lerner-Lam A. 2012. Advances in landslide nowcasting: evaluation of a global and regional modeling approach. *Environmental Earth Sciences* **66**: 1683–1696.

Lu N, Godt JW. 2013. *Hillslope, Hydrology and Stability*. Cambridge University Press: Cambridge, UK.

Mailier PJ, Stephenson DB, Ferro CAT, Hodges KI. 2006. Serial clustering of extratropical cyclones. *Monthly Weather Review* **134**: 2224–2240.

Malamud BD, Turcotte DL, Guzzetti F, Reichenbach P. 2004. Landslide inventories and their statistical properties. *Earth Surface Processes and Landforms* **29**: 687–711.

Marra F, Nikolopoulos EI, Creutin JD, Borga M. 2014. Radar rainfall estimation for the identification of debris-flow occurrence thresholds. *Journal of Hydrology* **519**(PB): 1607–1619.

Martelloni G, Segoni S, Lagomarsino D, Fanti R, Catani F. 2013. Snow accumulation/melting model (SAMM) for integrated use in regional scale landslide early warning systems. *Hydrology and Earth System Sciences* **17**: 1229–1240.

McCarthy M, Spillane S, Walsh S, Kendon M. 2016. The meteorology of the exceptional winter of 2015/2016 across the UK and Ireland. *Weather* **71**: 305–313.

Melillo M, Brunetti MT, Peruccacci S, Gariano SL, Guzzetti F. 2015. An algorithm for the objective reconstruction of rainfall events responsible for landslides. *Landslides* **12**: 311–320.

Met Office. 2003. 5 km resolution UK composite rainfall data from the Met Office Nimrod system. NCAS British Atmospheric Data Centre, Oxford. Available: http://catalogue.ceda.ac.uk/uuid/82adec1f896af6169112d09cc1174499 19 July 2017].

Meyer NK, Dyrrdal AV, Frauenfelder R, Etzelmuller B, Nadim F. 2012. Hydrometeorological threshold conditions for debris flow initiation in Norway. *Natural Hazards and Earth System Sciences* **12**: 3059–3073.

Meyer NK, Schwanghart W, Korup O, Nadim F. 2015. Roads at risk: traffic detours from debris flows in southern Norway. *Natural Hazards and Earth System Sciences* **15**: 985–995.

Milne FD, Werritty A, Davies MCR, Brown MJ. 2009. A recent debris flow event and implications for hazard management. *Quarterly Journal of Engineering Geology and Hydrogeology* **42**: 51–60.

Napolitano E, Fusco F, Baum RL, Godt JW, De Vita P. 2016. Effect of antecedent-hydrological conditions on rainfall triggering of debris flows in ash-fall pyroclastic mantled slopes of Campania (southern Italy). *Landslides* **13**: 967–983.

Nettleton I, Martin S, Hencher S, Moore R. 2005. *Debris Flow Types and Mechanisms*. Edinburgh: Scottish Government.

Parkes BL, Wetterhall F, Pappenberger F, He Y, Malamud BD, Cloke HL. 2013. Assessment of a 1-hour gridded precipitation dataset to drive a hydrological model: a case study of the summer 2007 floods in the Upper Severn, UK. *Hydrology Research* **44**: 89–105.

Pennington C, Harrison A. 2013. 2012: landslide year? *Geoscientist Online*. Available: http://www.geolsoc.org.uk/Geoscientist/Archive/June-2013/2012-Landslide-year [20 July 2017].

Pennington C, Dijkstra TA, Lark M, Dashwood C, Harrison A, Freeborough K. 2014. Antecedent precipitation as a potential proxy for landslide incidence in south west United Kingdom. In *Landslide Science for a Safer Geoenvironment. Vol. 1: The International Programme on Landslides (IPL)*, Sassa K, Canuti P, Yin Y (eds). Springer: Cham, Switzerland; 253–259.

Perkins NJ, Schisterman EF. 2006. The inconsistency of 'optimal' cut-points using two ROC based criteria. *American Journal of Epidemiology* **163**: 670–675.

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

Piciullo L, Gariano SL, Melillo M, Brunetti MT, Peruccacci S, Guzzetti F, Calvello M. 2016. Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. *Landslides* **14**: 995–108.

Postance B, Hillier J, Dijkstra TA, Dixon N. 2017. Extending natural hazard impacts: an assessment of landslide disruptions on a national road transportation network. *Environmental Research Letters* **12**: 14010.

Reid B. 2006. Global early warning systems for natural hazards: systematic and people-centred. *Philosophical Transations of the Royal Society A* **364**: 2167–2182.

Rota M, Antolini L. 2014. Finding the optimal cut-point for Gaussian and Gamma distributed biomarkers. *Computational Statistics and Data Analysis* **69**: 1–14.

Saito H, Nakayama D, Matsuyama H. 2010. Relationship between the initiation of a shallow landslide and rainfall intensity-duration thresholds in Japan. *Geomorphology* **118**: 167–175.

Schaefer JT. 1990. The critical success index as an indicator of warning skill. *Weather and Forecasting* **5**: 570–575.

Schellart ANA, Shepherd WJ, Saul AJ. 2012. Influence of rainfall estimation error and spatial variability on sewer flow prediction at a small urban scale. *Advances in Water Resources* **45**: 65–75.

Schisterman EF, Perkins NJ, Liu A, Bondell H. 2005. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology (Cambridge, MA)* **16**: 73–81.

Segoni S, Rosi A, Rossi G, Catani F, Casagli N. 2014a. Analysing the relationship between rainfalls and landslides to define a mosaic of triggering thresholds for regional-scale warning systems. *Natural Hazards and Earth System Sciences* **14**: 2637–2648.

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

Segoni S, Battistini A, Rossi G, Rosi A, Lagomarsino D, Catani F, Moretti S, Casagli N. 2015. Technical note: an operational landslide early warning system at regional scale based on space-time-variable rainfall thresholds. *Natural Hazards and Earth System Sciences* **15**: 853–861.

Sidle RC, Ochiai H. 2013. *Landslides: Processes, Prediction, and Land Use*. American Geophysical Union: Washington, DC.

Søren B, Graziella D, José C, Hervé C. 2014. Landslide thresholds at regional scale for an early warning system in Norway. In Proceedings of World Landslide Forum 3, Beijing; 2–6.

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.

Stoffel M, Tiranti D, Huggel C. 2014. Climate change impacts on mass movements: case studies from the European Alps. *Science of the Total Environment* **493**: 1255–1266.

Trewin NH. 2002. *The Geology of Scotland*, 4th rev. edn. Geological Society of London: London.

UNISDR. 2015. United Nations International Strategy for Disaster: Sendai Framework for Disaster Risk Reduction 2015–2030. UNISDR: Geneva.

Van Asch TWJ, Buma J, Van Beek LPH. 1999. A view on some hydrological triggering systems in landslides. *Geomorphology* **30**: 25–32.

Villarini G, Krajewski WF. 2008. Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges. *Advances in Water Resources* **31**: 1015–1023.

Villarini G, Mandapaka PV, Krajewski WF, Moore RJ. 2008. Rainfall and sampling uncertainties: a rain gauge perspective. *Journal of Geophysical Research-Atmospheres* **113** D11102: doi:https://doi.org/10.1029/2007JD009214.

Wachinger G, Renn O, Begg C, Kuhlicke C. 2013. The risk perception paradox: implications for governance and communication of natural hazards. *Risk Analysis* **33**: 1049–1065.

Waltham AC, Dixon N. 2000. Movement of the Mam Tor landslide, Derbyshire, UK. *Quartely Journal of Engineering Geology and Hydrogeology* **33**: 105–123.

Wieczorek GF. 1996. Landslide triggering mechanisms. In *Landslides: Investigation and Mitigation*. Transportation Research Board, National Research Council, special report. Turner AK, Schuster RL (eds). National Academy Press: Washington, DC; 673.

Wieczorek GF, Glade T. 2005. Climatic factors influencing occurrence of debris flows. In *Debris-Flow Hazards and Related Phenomena*. Springer: Berlin; 325–362.

Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*, 3rd edn. Cambridge, MA: Academic Press.

Wilson RC, Jayko AS. 1997. Preliminary maps showing rainfall thresholds for debris-flow activity, San Francisco Bay Region, California. US Geological Survey Open file report 97–745; 1–20.

Winter MG, Dent J, Macgregor F, Dempsey P, Motion A, Shackman L. 2010. Debris flow, rainfall and climate change in Scotland. *Quarterly Journal of Engineering Geology and Hydrogeology* **43**: 429–446.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article.

