Natural Hazards
and Earth System
Sciences

Predicting storm-triggered debris flow events: application to the 2009 Ionian Peloritan disaster (Sicily, Italy)

M. Cama[1], L. Lombardo[1,2], C. Conoscenti[1], V. Agnesi[1], and E. Rotigliano[1]

[1]Dipartimento di Scienze della Terra e del Mare, Università degli Studi di Palermo, Palermo, Italy
[2]Department of Physical Geography and GIS, University of Tübingen, Tübingen, Germany

Correspondence to: M. Cama (mariaelena.cama@unipa.it, mariaelenacama@gmail.com)

Abstract. The main assumption on which landslide susceptibility assessment by means of stochastic modelling lies is that the past is the key to the future. As a consequence, a stochastic model able to classify past known landslide events should be able to predict a future unknown scenario as well. However, storm-triggered multiple debris flow events in the Mediterranean region could pose some limits on the operative validity of such an expectation, as they are typically resultant of a randomness in time recurrence and magnitude and a great spatial variability, even at the scale of small catchments. This is the case for the 2007 and 2009 storm events, which recently hit north-eastern Sicily with different intensities, resulting in largely different disaster scenarios.

The study area is the small catchment of the Itala torrent ($10 \, \text{km}^2$), which drains from the southern Peloritani Mountains eastward to the Ionian Sea, in the territory of the Messina province (Sicily, Italy). Landslides have been mapped by integrating remote and field surveys, producing two event inventories which include 73 debris flows, activated in 2007, and 616 debris flows, triggered by the 2009 storm. Logistic regression was applied in order to obtain susceptibility models which utilize a set of predictors derived from a 2 m cell digital elevation model and a 1 : 50 000 scale geologic map. The research topic was explored by performing two types of validation procedures: self-validation, based on the random partition of each event inventory, and chrono-validation, based on the time partition of the landslide inventory. It was therefore possible to analyse and compare the performances both of the 2007 calibrated model in predicting the 2009 debris flows (forward chrono-validation), and vice versa of the 2009 calibrated model in predicting the 2007 debris flows (backward chrono-validation).

Both of the two predictions resulted in largely acceptable performances in terms of fitting, skill and reliability. However, a loss of performance and differences in the selected predictors arose between the self-validated and the chrono-validated models. These are interpreted as effects of the non-linearity in the domain of the trigger intensity of the relationships between predictors and slope response, as well as in terms of the different spatial paths of the two triggering storms at the catchment scale.

1 Introduction

Debris flows are among the most hazardous geological phenomena, which directly threat human lives in the light of their high energy and rapid propagation over slopes and drainage systems. In order to predict these phenomena, together with physically based approaches, which are mainly focused on the detection of the rainfall thresholds responsible for their triggering (e.g. Peres and Cancelliere, 2014; Bordoni et al., 2015) and on the physical modelling of the propagation phase (e.g. Schraml et al., 2015), susceptibility models (Brabb, 1984), suitable to depict prediction images of the sites where these phenomena are more likely to activate on a catchment/regional scale, are required as well. Combining the two approaches allows optimization of the use of early warning systems (e.g. Lagomarsino et al., 2015; Segoni et al., 2015; Stähli et al., 2015) – in doing so mitigating the debris flow risk on a catchment/regional scale.

Landslide susceptibility assessment can be achieved by means of different methods, among which the stochastic approach has gained increasing importance in the last 2 decades

in regional assessment applications. In fact, statistic models produce objective, quantitative and verifiable estimates of the spatial probability for new landslides in a given study area. Moreover, the stochastic approach is very easily implementable on geographic informative systems (GIS), making use of the very diffused nature of present databases of physical–environmental attribute layers. These methods are based on some generally accepted assumptions, the basic one being *the past is the key to the future* (Carrara et al., 1995). Therefore, a susceptibility model constructed to reproduce a past known landslide spatial distribution, will also be able to predict the future locations of new failures. In particular, for a given study area, statistical techniques allow the derivation and testing of the multivariate relationships between the spatial distributions of an inventory of landslides (the *known target pattern*) for significance as well as testing a set of physical–environmental variables (the predictors), which, acting as controlling factors, are supposed to drive the slope failures, on the basis of a geomorphological model. In the framework of the above-recalled principle, the new landslides (the outcomes) will occur under the same conditions which explain the known landslide distribution. Thus, a calibrated predictive model optimizes the functional relations between predictors and outcomes, maximizes its skill in fitting the known target pattern (the calibration data set), and it is finally tested for its correct reproduction of the unknown target pattern (the validation data set). As the controlling factors are selected among the time-invariant preparatory causes, regardless of how old the landslide inventory employed to calibrate the model is, as far as the basic assumption holds, any calibrated model will be able to predict any past or future unknown target pattern.

Unfortunately, very often, susceptibility assessment studies are affected by a lack of temporal information on the landslide inventory, which makes it impossible to perform a pure temporal or chrono-validation.

Based on the scheme described above, in order to elude the lack of temporal information, strategies for the validation of the predictive models can be defined. Specifically, when seasonal or event inventories (Guzzetti et al., 2012) are not available, a validation can be performed by following a *random time partition* procedure (Chung and Fabbri, 2003). In this case, the source inventory is split into a calibration and a validation subset to simulate the known and the unknown target patterns, respectively. In this work, the above scheme is defined as a self-validation procedure to highlight the notion that, under a morphodynamic perspective, calibration and validation patterns are actually two partial and complementary sides of the same event. Conversely, the term chrono-validation will be used when referring to pure temporal verification (Guzzetti et al., 2005), i.e. when the training and the test target patterns belong to two temporally separated data sets. A third scheme, frequently adopted for model spatial transferability or exportation (e.g. Von Ruette et al., 2011; Costanzo et al., 2012a; Lombardo et al., 2014; Petschko et

al., 2014), is based on the adoption of two different catchments or areas for calibration and validation (*spatial partition*).

It is evident how the whole scheme of the stochastic approach is strictly dependent on the basic assumption being held. Any changes in the real relationships between preparatory causes and landslide activity will affect the prediction skill of the obtained susceptibility models. Extreme events produce morphodynamic responses that can lie outside of the general rule. In fact, due to intense triggering, such as a storm, the same area can result in an "out-of-range" slope response because it could not be correctly predicted by a model skilled in fitting "normal" landslide scenarios. This could be a result of the non-linearity of the relationship between preparatory causes and landslides in the domain of the trigger intensity. Besides, the Mediterranean storms are typically affected by randomness in time recurrence and magnitude and a great spatial variability, even at the scale of small catchments. It is therefore necessary to check for this kind of behaviour to find a strategy which maximizes the ability of a susceptibility model to predict extreme events.

In spite of the wide diffusion of landslide susceptibility studies by means of statistical modelling, few cases are focused on detecting predictive limits when facing storm-triggered multiple debris flow events (e.g. Von Ruette et al., 2011; Tseng et al., 2015). In particular, the application of specific validation strategies to evaluate the effect of the trigger phenomena in modifying the predictive performance of the models is very rare. A contribution to this topic is presented here, drawing on a case study in north-eastern Sicily, where two recent storm events (2007 and 2009) hit the Ionian side of the Peloritani Mountains (Fig. 1) with different intensities. Specifically, the study area is the Itala catchment (nearly $10 \, \text{km}^2$), which is located in the southern sector of the Peloritan ridge.

In order to investigate our topic, the debris flows activated on the occasion of the two extreme events were mapped by integrating remote and field surveys, and a simple set of predictors was prepared by utilizing a 1 : 50 000 scale geological map and a 2 m cell digital elevation model (DEM). Statistical models were obtained by applying the stepwise (forward) binary logistic regression technique (Hosmer and Lemeshow, 2000), which has been largely adopted in landslide susceptibility studies (Atkinson et al., 1998; Ohlmacher and Davis, 2003; Süzen and Doyuran, 2004; Brenning, 2005; Carrara et al., 2008; Costanzo et al., 2014; Lombardo et al., 2014; Heckmann et al., 2014), demonstrating suitability for the geomorphological task and producing high performances, also in comparative studies (Guzzetti et al., 2006; Othman et al., 2015; Rossi et al., 2010). Multi-temporal high-resolution images (provided by ARTA – Assessorato Regionale Territorio e Ambiente) were made use of in order to prepare two landslide event inventories (Guzzetti et al., 2012), so that two types of modelling procedure are performed and validated: self-validation, based on the random partition into a
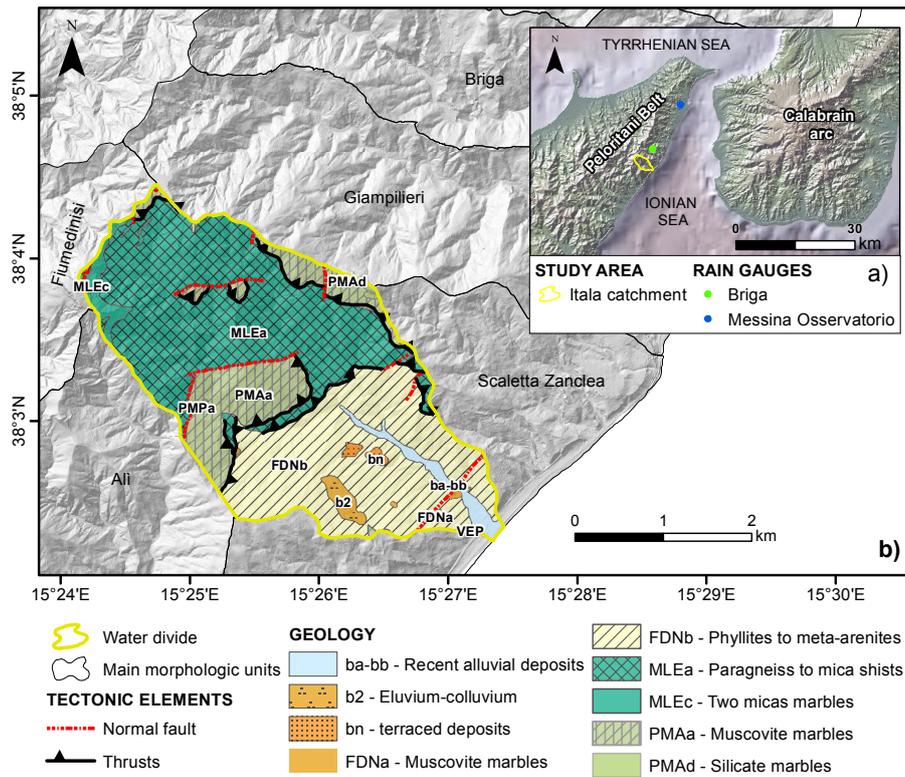
Figure 1. Study area: **(a)** geographical setting and rain gauge locations; **(b)** geology.

calibration and a validation subset of each event inventory, and chrono-validation, based on the temporal partition into the 2007 and 2009 cases. The latter procedure was applied to analyse the performances both of the 2007 calibrated model in predicting the 2009 debris flows source areas (forward chrono-validation) and of the 2009 calibrated model in predicting the 2007 debris flow source areas (backward chrono-validation). By analysing and comparing the predictive performances of binary logistic regression for the four types of models, the role of the triggering rainfall intensities is outlined and discussed.

2 Background

Testing a susceptibility model against future landslides is quite a hard task, especially because it would require researchers to "wait for the future to happen" (Guzzetti, 2005). Nevertheless, when a multi-temporal landslide inventory is available, the validation can be performed using a temporal criterion to separate calibration and validation data sets. Among others, Guzzetti et al. (2005) performed a "temporal verification procedure" which evaluates the effect of five landslide inventory updates on the performance of a susceptibility model. Similarly, other authors used a temporal criterion to validate the results of landslide susceptibility analysis at different scales (Zêzere et al., 2004; Vergari et al., 2011;

Wang et al., 2014), but none of them worked with storm-triggered debris flows and event inventories. Von Ruette et al. (2014) adopted a spatial partition scheme, with a partial insight into temporal validation which was limited for predicting the landslides triggered by two rainfall events in two close, but different, catchments. Chang et al. (2014) concentrated their focus on exploring the role of rainfall in controlling the chrono-validation performance for a much larger-scale case, demonstrated in a larger area (2868 km²), where a network of 24 rain gauges recorded nine great typhoon events.

The Messina area (Fig. 1) and the debris flow event of 2009 have been the focus of study in several scientific articles centred on different topics. Several studies have been devoted to the implementation of remote and semi-automatic techniques for landslide recognition and mapping of such a significant multiple occurring regional landslide event (Ardizzone et al., 2012; Mondini et al., 2011; Ciampalini et al., 2015). Del Ventisette et al. (2012) focused their research on the Giampilieri village area, analysing the triggering mechanism and estimating the volumes involved in the debris flow. They also applied a method based on conditional analysis in order to obtain a susceptibility map. Goswami et al. (2011) and De Guidi and Scudero (2013) explored the relationship between tectonic setting and landslide susceptibility, taking the Giampilieri and Scaletta catchments as study areas. Re-

ichenbach et al. (2014) evaluated the influence of land-use change on debris flow susceptibility for the Briga catchment. Stancanelli and Foti (2015) compared two different numerical models for simulating the 2009 debris flow event in the lower coastal sector of the affected area. Aronica et al. (2012a) published a detailed description of the 2009 event, with an insight into the saturation conditions of the soils and an evaluation of the difference of DEMs in the total volume of mobilized material for the Giampilieri catchment. Rainfall thresholds for the landslide activations have been investigated by Gariano et al. (2015), in the framework of a regional study, and by Peres and Cancelliere (2014), who conducted a specific study on the Ionian-Peloritan area, hit by the 2009 event. Lombardo et al. (2014) tested spatial exportation techniques for logistic regression-based susceptibility models, in the Briga and Giampilieri catchments.

With regards to the 2007 event, Aronica et al. (2012b) applied a physically based modelling tool to simulate the debris flows affecting a very small catchment, located 5 km south of the Itala stream.

In contrast to the above-mentioned research, in this paper, by studying two well split event inventories produced by two triggering events with different intensities, the relationship between controlling factors of triggers and morphodynamic responses are examined, and their effects on the predictive performance of stochastic susceptibility modelling are verified. Moreover, until now, no study has been published for the Itala catchment on the 2007 event, nor chrono-validated models and maps have been produced for the 2009 event.

3 General framework

3.1 Study area

The study area is located in the north-easternmost edge of Sicily (southern Italy), on the Ionian slopes of the Peloritan ridge, 20 km southward from the town of Messina (Fig. 1a). Specifically, the Itala catchment is located in the Itala municipality territory, stretching 10 km^2 and whose torrent drains south-eastward for near 6 km from Mt. Scuderi (1259 m a.s.l.) to the Ionian Sea. Geologically, the area is situated between the Mandanici, Mela and Aspromonte structural units (Messina et al., 2004), which are separated by thrusts and further fractured by the neo-tectonic faults. These units are made of high- to medium-grade metamorphic rocks. In particular, the Mandanici unit is primarily characterized by the outcropping of phyllites, while Mela and Aspromonte units mainly consist of paragneisses and mica schists (Fig. 1b).

According to the Köppen classification (Köppen, 1923), the climate in the region is classified as Mediterranean (Csa)-type, being therefore characterized by a dry season from April to September and a wet season from September to March, with an average yearly rainfall of nearly 900 mm. In
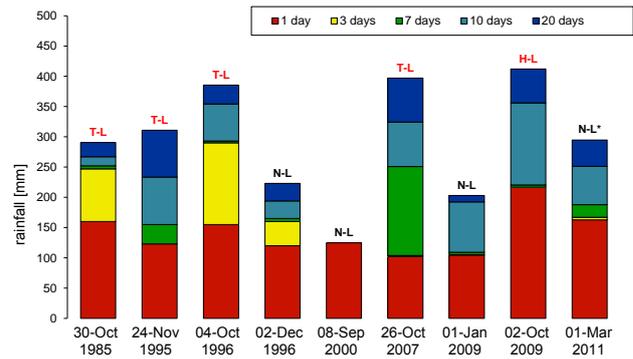
This event was responsible for the activation of tens of debris flows in a sector located about 5km South of Messina, but no landslides are reported for the Itala catchment.

Figure 2. Bar plot showing the cumulative (1 day, 3, 7, 10 and 20 days) rainfall in mm respectively for the main nine events recorded in the Itala catchment area.

addition, due to the warm water of the Mediterranean Sea and the proximity of the ridge to the sea coast, storm events are frequent in the autumn season in this area of Sicily.

Due to the limited length, together with high steepness of the Ionian-Peloritan torrents, although they are usually almost dry, under raining conditions the discharge can rapidly increase, causing floods which affect the infrastructure (especially roads) located in the proximity of the riverbanks. Moreover, during autumn storm events, the combination of the hydrologic regime and geomorphologic setting occasionally determines severe morphodynamic responses, including multiple debris flows and debris flood events, such as those which occurred in 2007 and 2009. The potential occurrence of this kind of event makes the whole area of Ionian-Peloritan catchments one of the most exposed zones to hydrogeological risk in Sicily.

The inhabited areas of the Itala catchment are located in very dangerous areas, either at the base of very steep terraced slopes, or near the outlet of the streams. With respect to the land use, the area can be divided into an eastern and a western sector. The former is highly terraced and mainly cultivated with citrus groves; the latter is characterized by chestnut forests and pastures. The study area is strongly affected by wildfires during the summer season; this influences the density of vegetation, the soil structure and the erosional processes acting on the slopes.

3.2 Historical records of rainfall events

The storm events of 2007 and 2009 have been analysed on the basis of two rain gauges belonging to the "Osservatorio delle Acque Sicilia", located in Briga and Messina Osservatorio (Fig. 1a). In particular, as the Peloritan area was historically hit by other storm events, a detailed analysis of antecedent rainfall conditions and of the historical record of debris flow events was carried out. The most important extreme meteorological events were selected first and, on the basis of the his-
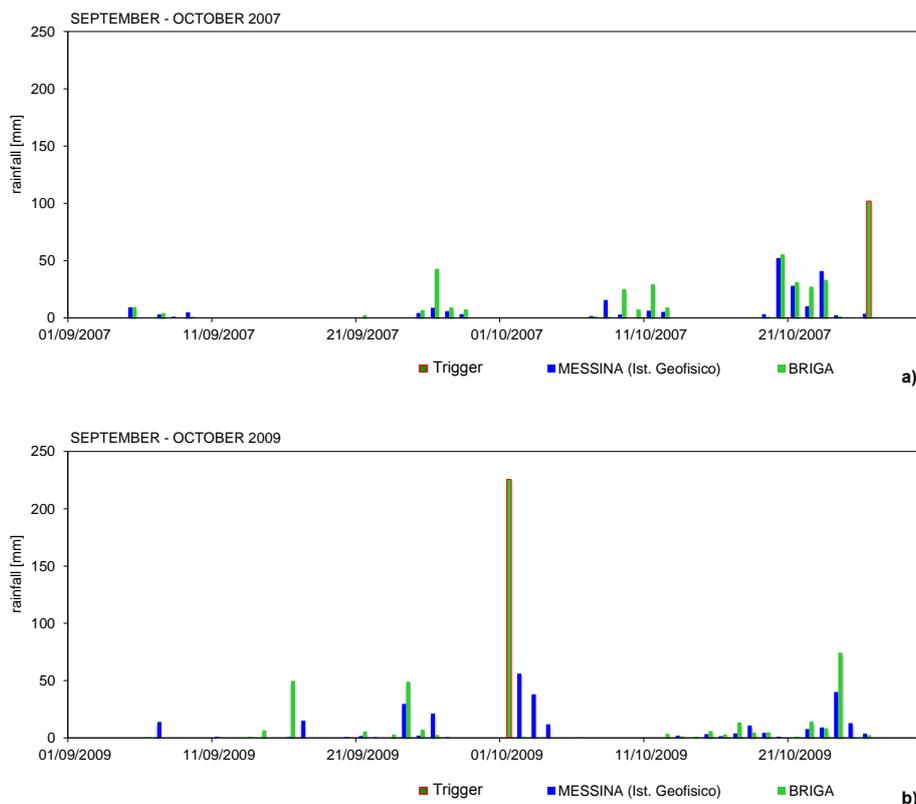
Figure 3. Time series of 2 months' precipitation for the Messina (Ist. Geofisico) and Briga rain gauges: **(a)** October 2007; **(b)** October 2009.

torical archive of the two main local newspapers ("Gazzetta del Sud" and "Giornale di Sicilia"), the associated landslide activity was identified. However, the estimation of the severity of the slope responses to the triggering storms cannot be accurately assessed at a basin scale from this kind of historical data. Therefore, with the exception of the 2007 and 2009 inventories, the classification of the debris flow events was limited to a qualitative ordinal scale (no landslides: N-L; tens of landslides: T-L; hundreds of landslides: H-L), based on the significance and frequency of damage reported for the Itala catchment area.

By analysing the daily cumulated rain from 1975 to 2011 (with the exception of 6 years with no rainfall data: 1987, 1988, 1989, 2003, 2004 and 2005), the nine heaviest rainfall events were detected on the basis of a 100 mm threshold, which corresponds approximately to the rain quantity recorded during the 2007 event. Figure 2 shows the 1-,3-, 7- and 20-day cumulated rainfall for the nine events, together with the corresponding debris flow activity reported for the Itala catchment area (indicated by red labels on the bar plot). Among the nine selected events, only five caused important multiple occurrence of debris flows, whose effects were reported in local newspapers. In fact, for the cases of 2 December 1996, 8 September 2000 and 20 January 2009, no landslide events were reported in local newspapers, which could indicate that either no landslides were activated or that

they were not significant enough in terms of damage caused to the villages. In these cases, the daily peak of rain was not anticipated by significant rain in the previous days. The more intense event of 1 March 2011 was responsible for the activation of tens of debris flows in a sector located about 5 km south of Messina, but no landslides are reported for the Itala catchment.

Among the events which caused reported landslides, the 30 October 1985 and 4 October 1996 events have very similar characteristics. In both cases, the main events were anticipated by significant precipitation in the antecedent 72 h. In contrast to this, the event of the 24 November 1995 was recorded with 123 mm day^{-1} and 155 mm week^{-1}. Looking at the 3 days and 7 days before the main event, the quantity of rain does not seem intense enough to lead to multiple debris flow occurrences, as it is very similar to the rate of precipitation recorded on 2 December 1996, when no landslides were triggered. Nevertheless, if a longer interval is considered (10 and 20 days) the cumulative quantity of rain exceeded 300 mm. This could justify the landslides being activated on this occasion, which were reported in the journal "Gazzetta del Sud" on 26 and 27 November.

The 26 October 2007 and the 1 October 2009 events are quite distinct when compared to the others. In fact, on the one hand, the 2007 daily rainfall event was anticipated by 3 dry days and heavy rainfall condition in a period of a week;
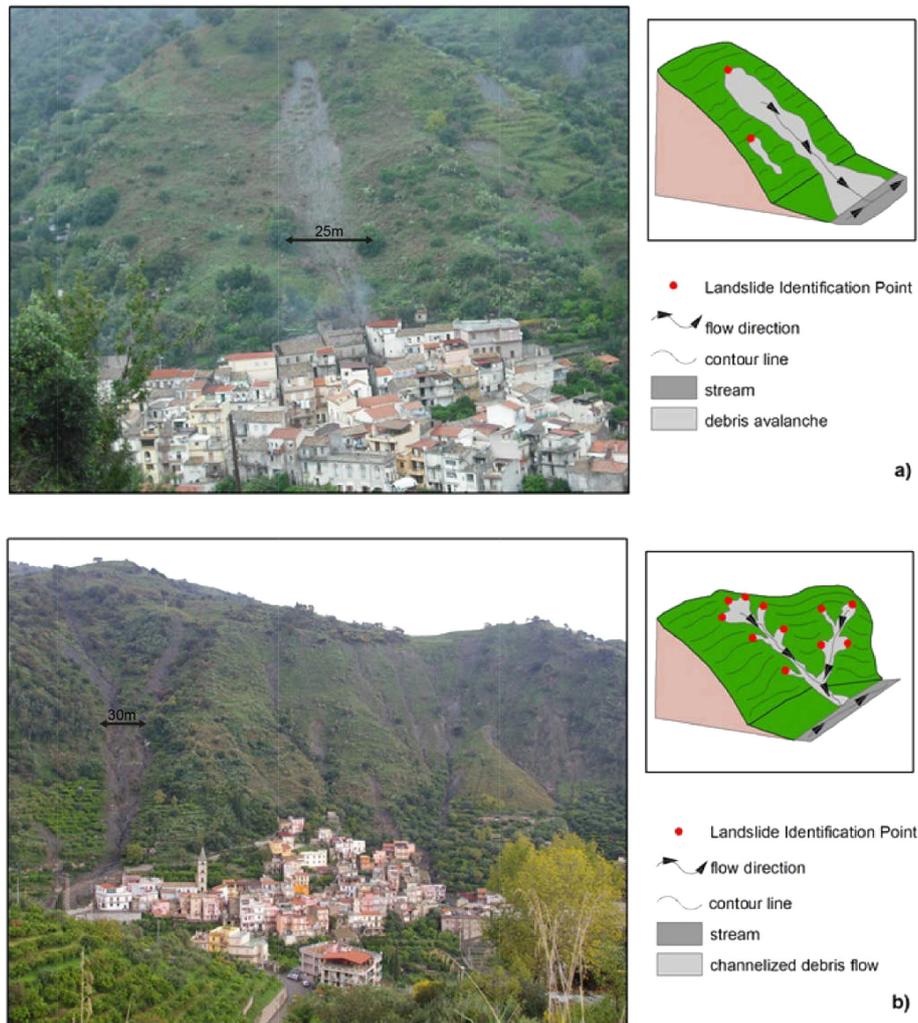
Figure 4. Overview of the area hit by the 2009 event: **(a)** Guidomandri village: debris avalanches are observable on the triangular facets parallel to the coast; **(b)** Itala village: channelized debris flows crossing the urbanized area.

on the other hand, the severity of the 2009 rainfall event is evident in both the daily (more than 200 mm) and in the 10- and 20-day precipitation, which exceeded 350 and 400 mm, respectively. In particular, Fig. 3a shows that the main event in 2007 (registered at Briga with 102 mm of rain in 24 h) was anticipated by longer and more extended raining periods, which lasted from 20 to 23 October, resulting in a cumulative weekly rainfall of 220.4 mm. The storm triggered hundreds of debris flows in the whole area, but only 73 in the Itala catchment. The 2009 event (Fig. 3b) presented the highest daily rain (nearly 220 mm); moreover, it followed two previous events: on 16 September, 49.2 mm in 6 h; and on 23–24 September, 79.6 mm in 10 h, determining a cumulative rain quantity which exceeded 412 mm in 20 days. As a consequence, on 1 October 2009 in an area of less than 10 km^2, hundreds of debris flows and debris flood events caused large damage to buildings and main roads in the Itala catchment.

To give a view of the large spatial variability of rainfall storms in this area, it is worth noting that although intense rainfall was recorded at the Briga rain gauge (102 and 220 mm day^{-1}, respectively), low values were recorded for the Messina Osservatorio (3.6 mm day^{-1} and no rain, respectively). This demonstrates that such extreme events are very localized, with rainfall conditions significantly changing in a range of a distance of only 15 km. However, although the authors believe that the small-scale rainfall distribution is very important for the prediction of the debris flow locations, the rain gauge network is not dense enough to evaluate the variability of the rain conditions at the catchment scale. Therefore, this variable cannot be introduced in the susceptibility models.

4 Materials and methods

The application of binary logistic regression (BLR) for landslide susceptibility assessment typically requires the following steps: the partition of the study area into mapping units, which are then characterized with respect to a set of potential predictors; the assignment of stability conditions to each mapping unit, based on its spatial relation with a set of known landslides (e.g. inclusion or intersection); the extraction of a balanced (stable/unstable) data set from the whole set of mapping units; the regression of the modelling function; and the verification of the performance of the model in correctly predicting stability conditions for each pixel, the latter defined on the basis of a set of unknown landslides.

This chapter describes the methods and the model building strategies which have been adopted to investigate the main research topic: exploring skills and limits in predicting the source areas of storm-triggered debris flows.

4.1 Landslide inventory

The typologies of the landslides that were activated during the 2007 and 2009 events are mainly classified as channelized debris flows and debris avalanches or hillslope debris flows (Varnes, 1978; Hutchinson, 1988; Hungr et al., 2001, 2014), which affected the weathered mantle of the metamorphic bedrock on the very steep slopes of the Itala catchment (Fig. 4). However, as this paper aimed to study susceptibility to new activations or the prediction of source areas, the whole set of phenomena was processed as a single type, using in the following the general sense of the term *debris flow*. The very few cases of bedrock landslides, such as falls and rotational slides, were deliberately excluded from the analysis, as they would have required a different approach both in terms of controlling factors and statistical methods.

Landslide recognition was performed by integrating a field survey, which was carried out soon after the 2009 disaster, and orthophoto analysis which allowed the slopes to be visualized at different dates. In particular, high-resolution lidar (Light Detention And Ranging) data were used from two different acquisitions, 2008 and 2009, respectively. These data were provided by the Territory and Environment Department of the Sicilian government (ARTA 2008 – Assessorato Regionale Territorio e Ambiente) and the National Civil Protection (PCN 2009, Protezione Civile Nazionale). The ARTA 2008 data (taken in August) include 0.25 m pixel orthophotos and a DEM, with 2 and 0.22 m for horizontal and vertical resolution, respectively. The PCN 2009 data were acquired 6 days after the 2009 event and includes 15 cm pixel orthophotos, and a 1.1 m cell DEM. In addition, multi-temporal (2005, 2006, 2010 and 2012) Google Earth™ (GE) images were analysed in order to compare the 2007 and 2009 mapped phenomena with the previous and following slope conditions.

An event inventory (Guzzetti et al., 2012) has to report only those landslides which have been triggered by a single specific trigger occurrence, such as an earthquake, rainfall or snowmelt. To fit this constraint, first landslide mapping was carried out on the 2008 and the 2009 images, obtaining a first version of the 2007 and 2009 inventories. However, the mapped landslides were supposed to be activated during 26 October 2007 for the first inventory and 1 October 2009 for the second. Therefore, the morphologies mapped on 2007 were also compared with the 2006 GE images. By combining the data obtained from the three time frames, five different cases were obtained (Fig. 5): (a) debris flows mapped on the 2007 orthophotos but which activated before the 2007 event; (b) debris flows which activated during the 2007 event but did not reactivate or retreat during the 2009 event; (c) debris flows which activated during the 2007 event that retreated or reactivated during the 2009 event; (d) debris flows which activated during the 2007 event which had been completely eroded during the propagation phase of the 2009 event and (e) debris flows which activated during the 2009 event in precedent stable areas.

The final event inventories (Fig. 6) contained 73 debris flows for 2007, corresponding to cases (b), (c) and (d), and 616 for 2009, corresponding to case (e). Each landslide inventory was stored in two separated vector layers: the first containing a polygon representing the source areas, and the second containing the landslide identification points (LIP), corresponding to the highest point along the crown of each mapped phenomenon (Costanzo et al., 2012b, 2014; Lombardo et al., 2014).

4.2 Binary logistic regression

Binary logistic regression (BLR) is a multivariate statistical technique, based on a frequentist approach, which is used to model the expected value of a response variable (the outcome) by a linear combination of either continuous and/or discrete predictor variables (Hosmer and Lemeshow, 2000). With respect to other frequentist methods (e.g. discriminant analysis), it does not require any linearization or transformation to obtain normal distributed covariates. Moreover, the outcome of BLR is easily interpretable for applied scientists.

In binary logistic regression the response variable Y assumes one of the two mutually exclusive values of 0 (no landslide) or 1 (landslide) for stable mapping units or unstable mapping units, respectively.

The relationship between the predictors and the probability for the response variable to assume the value 1 is linearized by the logit function (Y), which corresponds to the following transformation:

$$\text{logit}(Y) = \ln[P(Y=1) / (1 - P(Y=1))]$$
$$= \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n, \tag{1}$$

where $P(Y=1)$ is the probability that the response variables assumes the value 1, α a constant term or intercept, x_1, x_2, $\ldots x_n$ are the input predictor variables and the β_n their coefficients. Therefore, once the logit function is calculated,
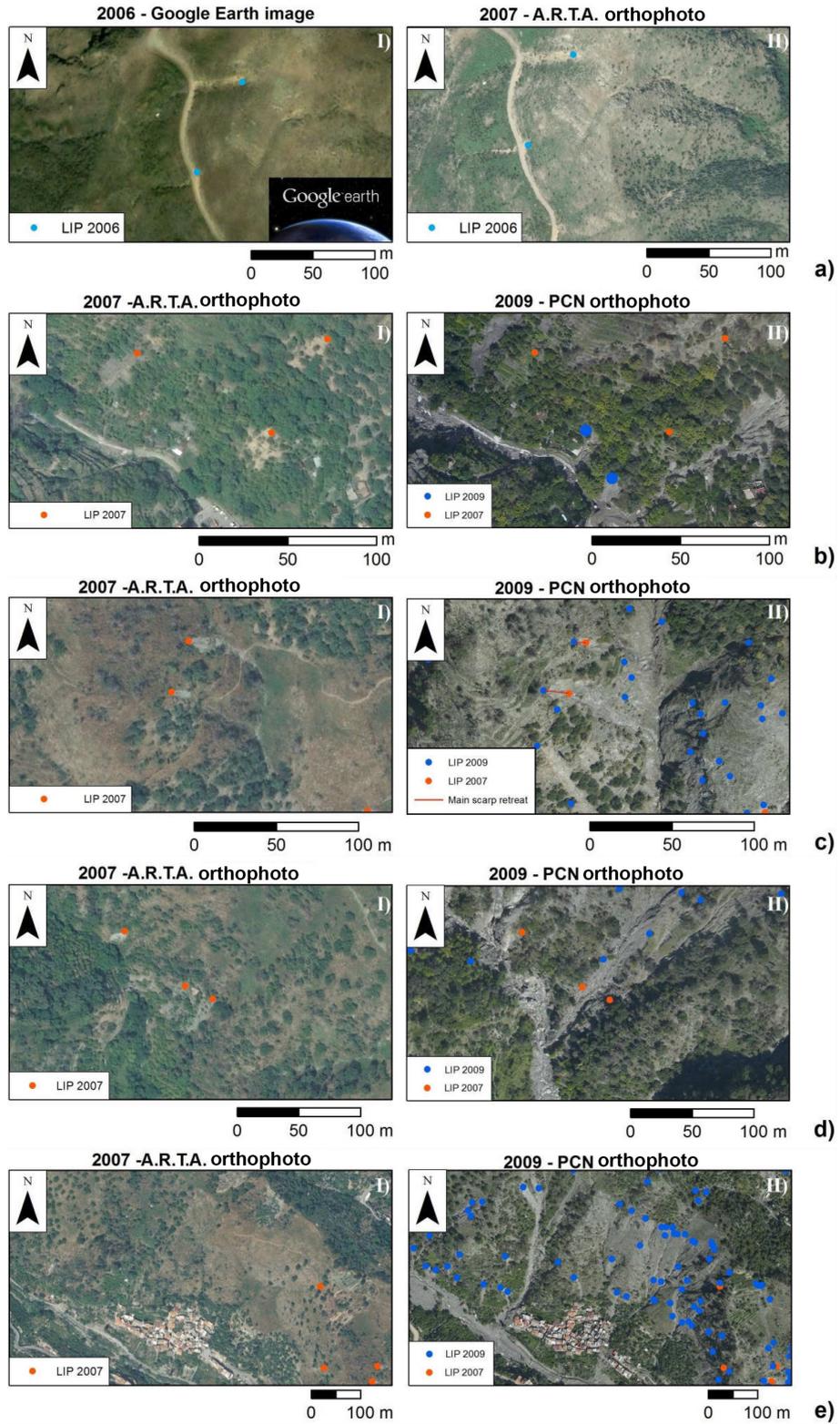
Figure 5. Comparison of morphologies between two different images resulting in five different cases: **(a)** debris flows recognized on the 2007 orthophoto but which activated before the 2007 event; **(b)** debris flows which activated in 2007 which did not reactivate or retreat in 2009; **(c)** debris flows which activated in 2007 that retreated or reactivated in 2009; **(d)** debris flows which activated in 2007 which were completely included in 2009; **(e)** debris flows which activated in 2009.
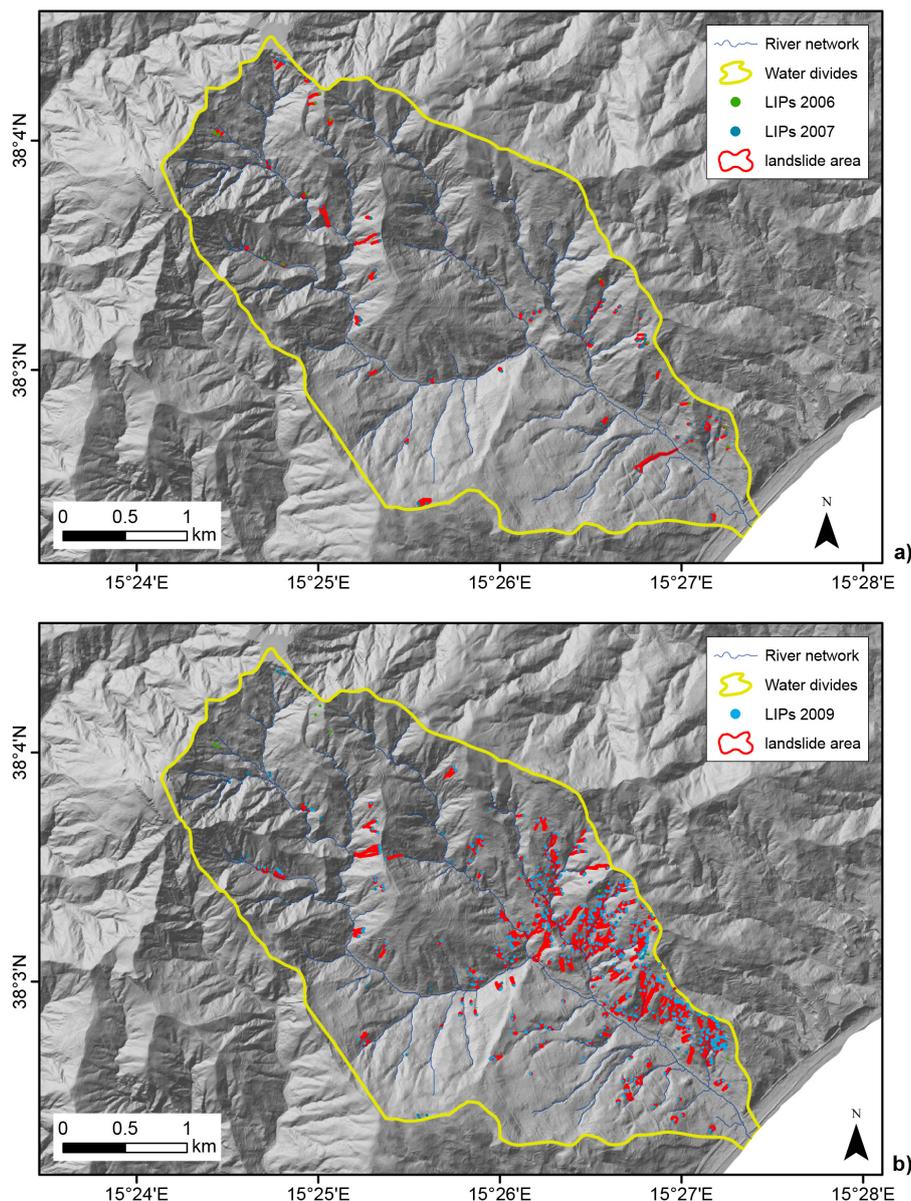
Figure 6. Debris flow event inventories: **(a)** 2007 inventory containing 73 debris flows; **(b)** 2009 inventory containing 616 debris flows.

and the $\beta_1, \beta_2, \ldots \beta_n$ values are known, the probability can be back-calculated using the following formula:

$$P(Y = 1) = e^{\text{logit}(Y)} / [1 + e^{\text{logit}(Y)}]. \qquad (2)$$

This equation ensures that, for any given case, the probability $P(Y = 1)$ will not be less than 0 or greater than 1 with logit (Y) ranging in the full $\pm \infty$ interval.

The odds ratios (OR), which are calculated by simply exponentiating β_n, indicates how likely (or unlikely) it is for the outcome to be positive (unstable cell) when a unit change of an independent variable occurs (Hosmer and Lemeshow, 2000). Negatively correlated variables will produce negative β_n and OR limited between 0 and 1; positively correlated variables will result in positive β_n and OR greater than 1.

In order to estimate the best intercept and β_n coefficients, the logistic regression uses the maximum likelihood algorithm. This maximizes the value of the log-likelihood function (LL), which indicates how likely is to obtain the observed value of Y, given the values of independent variables and coefficients (Menard, 2002). In particular, the global fitting of the regressed model on the data domain is usually expressed by $-2LL$ (negative log-likelihood) which is an estimator based on the maximum likelihood criterion. The differences in $-2LL$ value between the model with only the intercept ($L_{\text{INTERCEPT}}$) and the full model (L_{MODEL}) have a χ^2 distribution, so that the significance of the regressed coefficients can be easily tested (Ohlmacher and Davis, 2003; Ak-
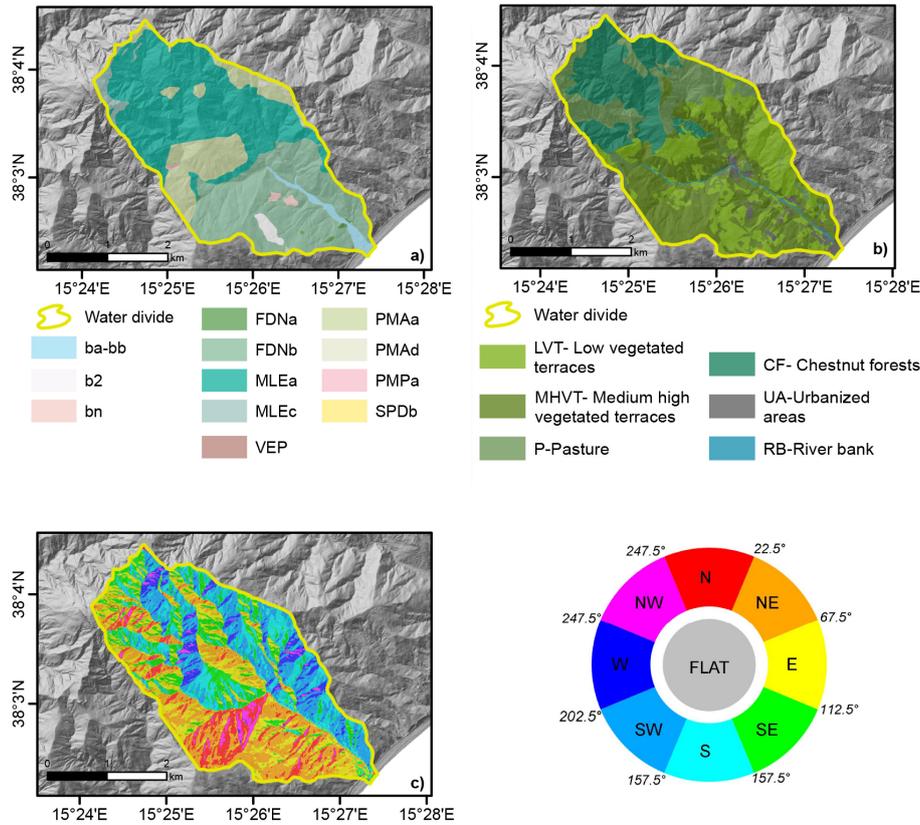
Figure 7. Discrete variables: **(a)** outcropping lithology (GEO; see Fig. 1 for description); **(b)** land use (USE); **(c)** aspect (ASP).

gun and Turk, 2011). In other words, the −2LL test estimates the significance of the increase in model fitting produced by the introduction of the predictors.

In the present research, we applied BLR under a stepwise selection routine, which has already been successfully adopted in landslides and debris flow susceptibility studies (Begueria, 2006; Meusburger and Alewell, 2009; Atkinson and Massari, 2011; Costanzo et al., 2014; Heckmann et al., 2014; Lombardo et al., 2014). The stepwise selection is an iterative procedure, which selects the best performing and most parsimonious set of predicting variables. It can be performed either in forward or in backward mode. In the first case, the procedure starts from an "intercept only" model and consists in selecting and adding, at each step, the variable which maximises the log-likelihood value. On the contrary, the backward stepwise selection starts from a full model, including all the variables, and removes the variables iteratively until the model reaches the best fitting. In the forward stepwise selection, at every step the procedure introduces all the variables iteratively and selects the one that maximizes the −2LL values. The first factor to be included is the one that produces the greatest change in the log-likelihood, with respect to the intercept. Applying the chi-square distribution of the −2LL values, the iterative calculation stops when the significance level of the increase, produced by including a

new predictor, is lower than 1%. Thus, the final result is the restricted list of variables, each with its order of importance (i.e. the iteration in which it was picked up) that can be submitted to the final BLR.

All the statistical analyses which are hereafter discussed were performed by using open source software (TANAGRA: Rakotomalala, 2005).

4.3 Covariates and outcome status assignment

The first step in modelling the debris flow susceptibility using a stochastic approach is to select those mapping units in which the study area has to be partitioned. Mapping units are the basic spatial elements in which the model will be able to produce a prediction. Two main types of mapping units are adopted in literature: hydro-geomorphological units and regular grids. The former allows the model to take advantage of the morphodynamic homogeneity of the area which is included in each single unit, corresponding to hydrological or slope units; the latter optimizes the matching between the spatial resolution of the source layers of some important predictors, typically having the same grid structure of the DEM.

In the present research, a raster-based structure was adopted by partitioning the study area into a grid of 8 m
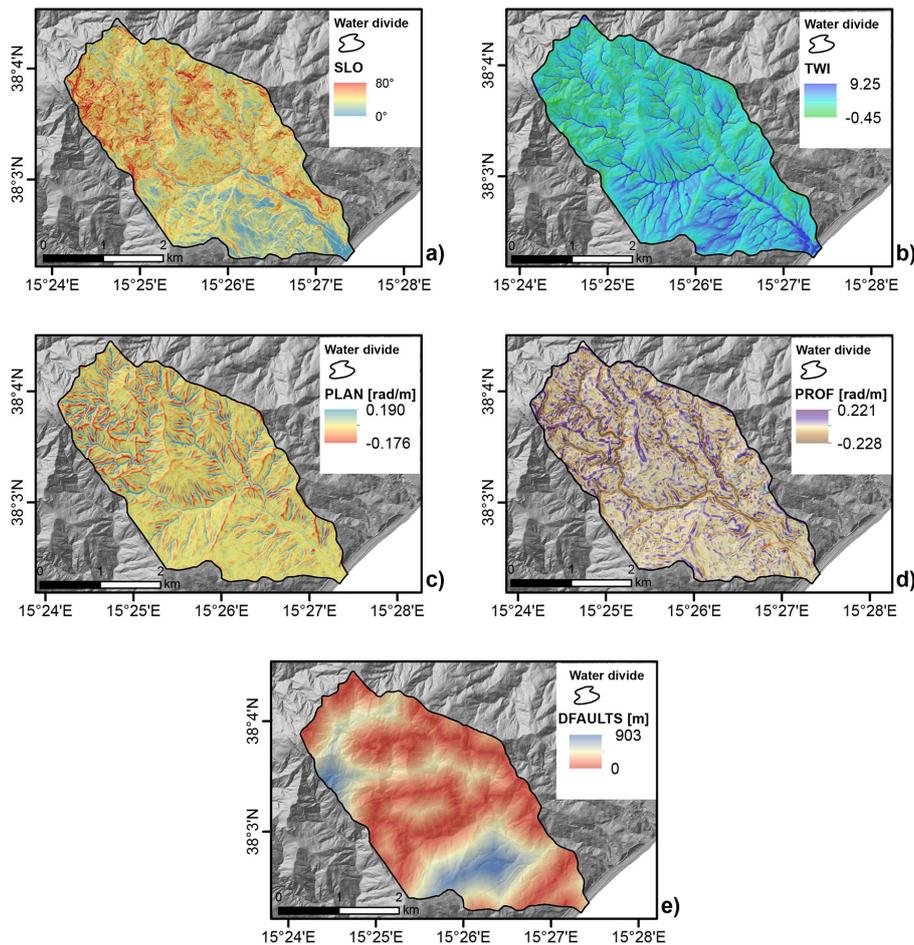
Figure 8. Continuous variables: **(a)** slope (SLO); **(b)** topographic wetness index (TWI); **(c)** plan curvature (PLAN); **(d)** profile curvature (PROF); **(e)** distance from tectonic elements (DFAULTS).

square cells, which required also the rasterization of the spatial distribution of all the covariates.

Starting from a DEM and a geological map, the following eight potential predictors have been selected and their value assigned to each cell in which the study area has been partitioned (Figs. 7 and 8): outcropping lithology (GEO), land use (USE), aspect (ASP), steepness (SLO), topographic wetness index (TWI), plan (PLAN) and profile (PROF) curvatures and distance from tectonic features (DFAULT).

Outcropping lithology and tectonic features are proxy variables expressing the mechanical properties of the bedrock and the weathered mantle. These variables were obtained from a 1 : 50 000 available geological map (Lentini et al., 2007), which was derived from 1 : 10 000 field surveys.

Information on land use allows the model to summarize those potential modifications of the natural structure of the regolithic mantle and the bedrock which are related to anthropogenic activities. In order to express these properties, a land-use map, based on the analysis of the orthophotos ARTA 2007/2008 and PCN 2009 and field recognition,

was prepared. The final land-use map contains six classes: (i) medium-high vegetated terraces (MHVT); (ii) low vegetated terraces (LVT); (iii) chestnut forests (CF); (iv) pastures (P); (v) urbanized areas (UA); (vi) river beds and beaches (RB).

Slope steepness, plan and profile curvatures are related to the energy of the relief. Steepness is commonly used as a predictor in landslide susceptibility and very often it demonstrates a very high importance. In fact, especially for debris flow analysis it is expected to be one of the most significant variables because it is directly linked to the shear strength acting onto the potential shallow failure surface. Moreover, for shallow failures presenting slide or flow mechanisms, the topographic surface and the rupture plane or zone can be considered as almost parallel. In this case, the slope steepness is a proxy for the real inclination of the potential failure surface. Steepness also controls the overland and subsurface flow velocity and runoff rate. At the same time, the topographic curvatures control the divergence and convergence, both of surface runoff and shallow gravitational stresses (Ohlmacher,

2007). Curvatures are expected to be the best proxy variables for convergent flow of water (plan curvature) and changes in flow velocity (profile curvature). In this study the *profile curvature* and the *plan curvature* were used, which correspond to the second derivatives of the slope steepness and the aspect, respectively.

The topographic wetness index is defined as $\ln(As / \tan\beta)$, where As is the local upslope area draining per contour unit length, and β is the local slope angle. It describes the extension and distribution of the saturation zones assuming steady-state conditions and uniform soil properties. By comparing the field data, it has been demonstrated that TWI can be considered a proxy variable directly related with the properties of soil, in particular with the soil moisture, A horizon depth, phosphorus content and organic matter (Moore et al., 1993).

Aspect controls the intensity of the solar insolation at the Earth's surface, and as a consequence, also the evapotranspiration and flora and fauna distribution and abundance. It is very important to consider the erosional processes related to the chemical physical weathering, operated by water, temperature and vegetation, in the determination of landslide susceptibility. Further, ASP frequently assumes a role of proxy variable for the attitude of the rock layers.

The source for the calculation of the topographic attributes was the DEM ARTA 2007/2008 subsequently resampled at 8 m pixel size with the nearest neighbour approach. The resampling operation on the original DEM (2 m pixel size) smoothed the effects of microtopography and possible noise existing on the original data.

All the factors were calculated using SAGA GIS (System for Automated Geoscientific Analysis, Conrad, 2007).

Once the layers of the predictors were obtained, they were combined in a multivariate grid whose cells status (stable/unstable) was defined on the basis of the intersection with the LIPs. Each cell hosting at least one LIP was set as unstable, in order to calibrate the models in predicting the locations of future LIPs, which in our scheme correspond to debris flow initiation areas.

4.4 Validation procedures and model building strategy

Model validation is a mandatory component of susceptibility assessment studies (Carrara et al., 2003; Guzzetti et al., 2006; Frattini et al., 2010; Rossi et al., 2010). No matter the method adopted in modelling the susceptibility, rigorous and quantitative validation procedures are the only criterion for accepting or rejecting a predictive model.

The validation of a model requires the availability of a calibration and a validation set of landslides or outcomes. The training landslides are applied to calibrate the maximum-likelihood fitting, so that the regression coefficients are optimized; the predicted probability which is generated by the model is then compared to the actual unknown target pattern which is defined by the validation landslides set. The accuracy of a model is then evaluated by comparing the produced prediction image to the known (calibration) and unknown (validation) target patterns. In particular, the degree of fit expresses the ability of the model to classify the known cases, while the prediction skill is the ability to predict the unknown cases.

As proposed by Chung and Fabbri (2003), calibration and validation data sets can be obtained by time partition, random time partition or spatial partition. The first is possible when multi-temporal landslides inventories are available, the second is based on randomly partitioning single-epoch data sets and the third on sub-dividing the study area into two similar sub-sectors. Random time partition procedures can be applied either on the landslide inventory (Conoscenti et al., 2008a) or on the mapping units database (Conoscenti et al., 2008b), whilst spatial partition can also be performed also on not nested or adjacent areas such as in the study aimed at susceptibility model exportations (von Ruette et al., 2011; Costanzo et al., 2012a; Lombardo et al., 2014).

However, validating a model requires precision, robustness and geomorphological adequacy or coherence for testing its accuracy, both in terms of predictive performance and inner structure of the model. The latter corresponds, in a stepwise BLR procedure, to the rank and the coefficients of the selected predictors (Frattini et al., 2010; Costanzo et al., 2014; Lombardo et al., 2014). Moreover, as BLR does for balanced (positive/negative cases) data sets, a single regressed data set must contain the positive cases (unstable cells) and an equal number of randomly selected negatives (Atkinson et al., 1998; Süzen and Doyuran, 2004; Nefeslioglu et al., 2008; Bai et al., 2009; Van Den Eeckhaut et al., 2009; Frattini et al., 2010; Costanzo et al., 2014), which could determine a low representativeness of the analysed cases. In particular, in this study, each pixel containing a LIP has been considered as being in the diagnostic area (Rotigliano et al., 2011), while the negative cases have been randomly selected in the catchment, outside the landslide polygons. In order to obtain a better dispersion of points and to avoid autocorrelation of the spatial variables, the distance in the random selection was maximized. Therefore, every model was composed of 146 balanced cases (positive/negative), for 2007, and 1232 balanced cases, for 2009. This heavily reduces the number of actually analysed cases to a very small percentage of the cells in which the study area is partitioned, so that a need of testing the representativeness of the worked subset also arises. To control the possible effects introduced by this procedure, multi-extraction of negatives are to be performed and more than one data set regressed. Specifically, a multiple extraction produces m different balanced data sets, each composed by the union of the same positives and a different set of randomly extracted negatives. Multi-fold cross validation procedures are then applied, by resampling the same data set n times to perform n replicates of the regression procedure, finally obtaining $n \times m$ outcomes of the same performance indexes or model parameters.

Table 1. Value prediction and confusion matrix of cross-folded validation for the 2007 data set.

Error rate			Mean	0.336	SD	0.028
Value prediction			Confusion matrix			
				Observed		
Value	Recall	1-precision		Yes	No	Sum
Yes	0.645	0.331	Predicted — Yes	1348	668	2016
No	0.683	0.340	No	742	1442	2184
			Sum	2090	2110	4200

Table 2. Value prediction and confusion matrix of cross-folded validation for the 2009 data set.

Error rate			Mean	0.219	SD	0.011
Value prediction			Confusion matrix			
				Observed		
Value	Recall	1-precision		Yes	No	Sum
Yes	0.777	0.216	Predicted — Yes	14 335	3948	18 283
No	0.786	0.221	No	4115	14 502	18 617
			Sum	18 450	18 450	36 900

In this research, two suites of 10 data sets were extracted for both the 2007 and 2009 models; a 10-fold cross validation procedure was then applied to each data set, which gave a total of one hundred probability estimates (10 replicates × 10 subsets) for each mapping unit, on which tests of accuracy and precision of the predictive performance were based. Moreover, each of the one hundred replicates resulted in a set of ranked predictors and regression coefficients, the comparison of which allowed us to test the precision and the robustness of the model.

Once a cut off for the estimated probability is fixed to split positive and negative predictions, the crossing with a target pattern results in the production of true positives (TP), true negatives (TN), false positives (FP: type I errors) and false negatives (FN: type II errors) cases. Contingency tables are used to summarize these data and to compute the model error rate, $(TP + TN) / (TP + TN + FP + FN)$, sensitivity or true positive rate, $(TP/(TP + FN))$, and $1 -$ specificity or false positive rate, $(FP / (TN + FP))$. Moreover, in order to assess the prediction accuracy of the models, the Hanssen and Kuipers (1965) (HK) skill score was also used. This index is defined as the difference between true positive and false positive. The HK maximum values measure the ability of the forecast system to discriminate between events and non-events. Maximizing these values means minimizing the probability range where the user would be unsure of the forecast.

A cut-off independent technique for estimating the accuracy of a predictive model is represented by receiver operating characteristic (ROC) curves, which depict the trade-off between success and failures for the decreasing probability threshold, in sensitivity versus 1-specificity plots. The area under the curve (AUC) in the ROC plots is the most adopted metric for the accuracy of the predictive models.

The precision and accuracy of the model can also be represented in spatial terms, by preparing prediction and error maps. For each mapping unit, the mean susceptibility and the dispersion of its estimates are plotted and compared to the actual distribution of the unknown positives.

In order to investigate the main research topic, two kinds of modelling procedures have been conducted. A self-validation scheme was applied for each of the two event inventories (2007 and 2009), by randomly splitting (90/10 %) the 10 extracted balanced data sets of the two temporal suites into a calibration and a validation subset. For each data set, the random splitting procedure was applied 10 times, resulting in one hundred self-validated replicates.

A chrono-validation scheme was then applied, by calibrating the model with the whole event inventory of each epoch and validating the performance in matching the event inventories of the other. We hereafter refer to forward chrono-validation, if calibrating with 2007 and validating with 2009, and vice versa to backward chrono-validation, if calibrating with 2009 and validating with 2007. For each temporal model suite, we produced 10 prediction images based on
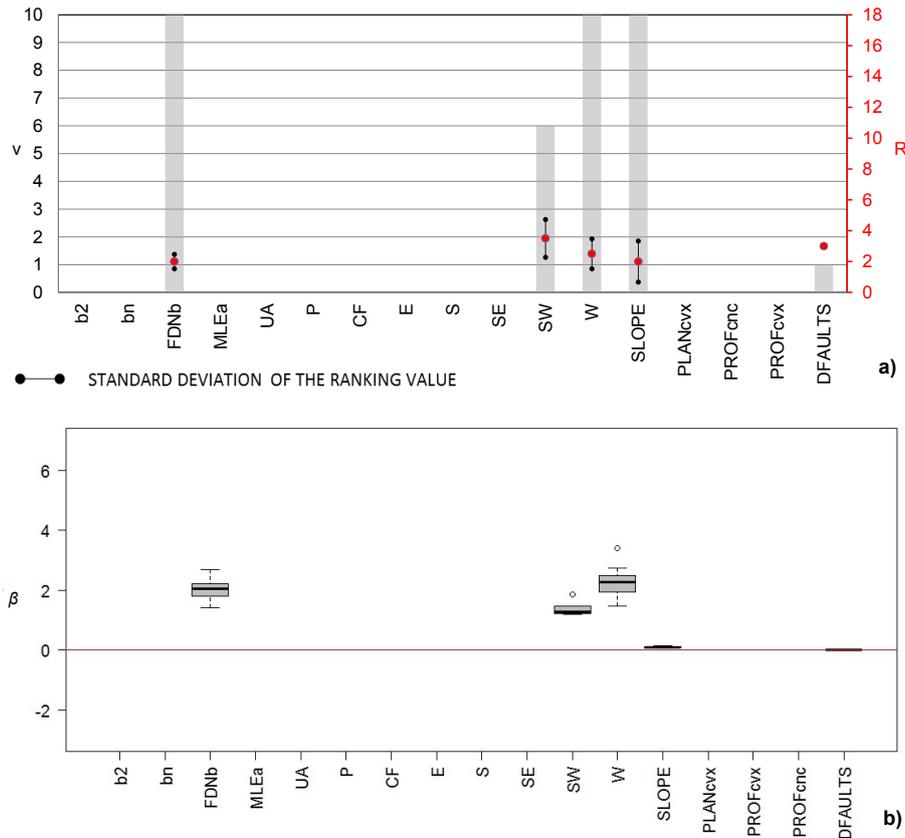
Figure 9. Selected variables for the 2007 suite of models: **(a)** ranking and frequency; **(b)** β values.

the 10 data sets of the other suite, again having one hundred backward and one hundred forward chrono-validated replicates.

5 Results

The results of the cross-validation procedures for the one hundred 2009 and 2007 self-validated models are presented in Tables 1 and 2. Generally, the 2009 models (Table 2) resulted in a better performing prediction with lower (0.336, for 2007; 0.219, for 2009) and more stable error rates. Similarly, the ROC-AUCs (Table 3) attested to the good quality of the models, with a higher performance for the 2009 model (2009 AUC was 0.85, 2007 AUC was 0.70) and no evidence of overfitting.

With regards to the predictors, the 2007 model suite selected five variables (Fig. 9), four of which had a frequency of more than 5/10: west and south-west slope aspect, steepness and phyllites to meta-arenites (FDNb) outcropping lithology resulted as the main causative factors for the 2007 debris flows. A larger set of variables (17) was included by BLR in the 2009 model suite (Fig. 10), 15 of which were selected more than five times. Among the topographic variables, the most important were: steepness, all the pixels without any

northward aspect component, profile curvatures (both concave and convex) and plan convex curvature of slopes. Together with topographic variables, FDNb and paragneiss to mica shists (MLEa) lithologies, distance from tectonic elements (DFAULTS) and chestnut forests (CF) and pastures (P) land-use classes were always selected with high and stable rankings. Concerning the β-coefficients, only profile curvature concavity, the variables DFAULT and CF and P land uses showed negative values, indicating inverse correlation with the debris flow source areas.

Once the overall quality of the predictive performance of the 2007 and 2009 models was assessed, regressions were run for the 10 full (without splitting into calibration and validation subsets) data sets of each event inventory, which maximized the fitting of the models. For both these full self-validated models (Fig. 11), the obtained ROC-AUCs are above the good performing threshold (> 0.81 for 2007; > 0.87 for 2009), with average error rates of 0.26 for 2007, and 0.22 for 2009. The 2007 and 2009 full models were then submitted to forward and backward chrono-validation, respectively, resulting in largely acceptable ROC-AUCs (> 0.75) and error rates (< 0.3), although a loss in the predictive performance of both the temporal predictions was observed. In particular, by comparing the self-validation and the chrono-validation per-
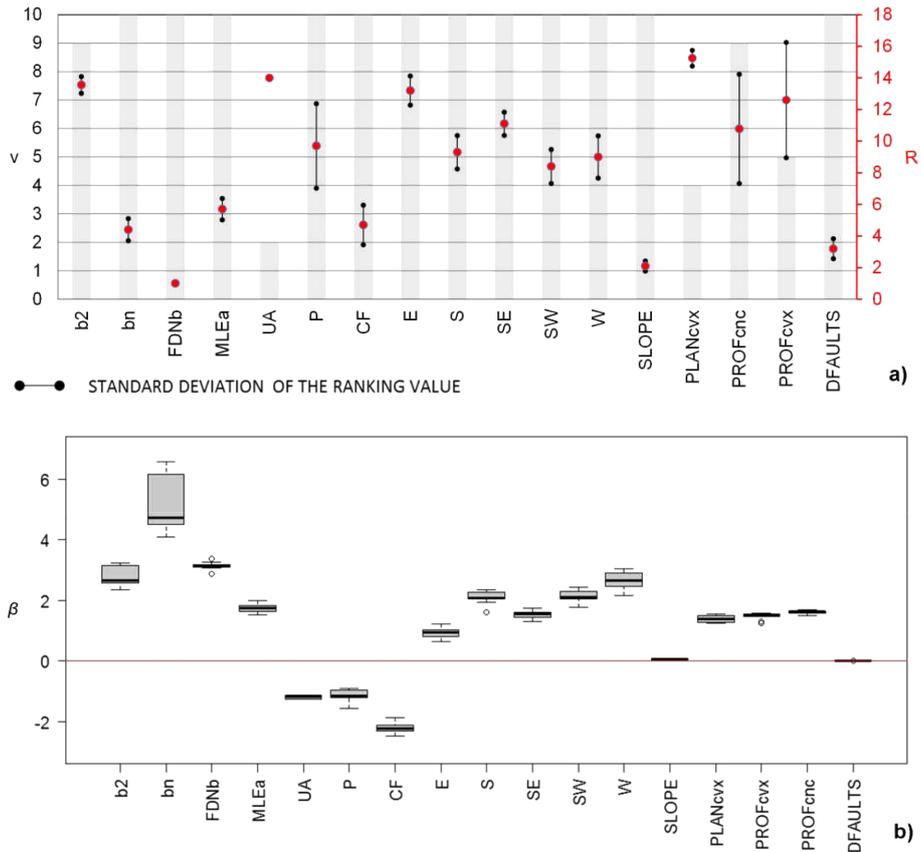
Figure 10. Selected variables for the 2009 suite of models: **(a)** ranking and frequency; **(b)** β values. For purposes of representation, the coefficients of the topographic curvatures are reported as $\log \beta$ values.
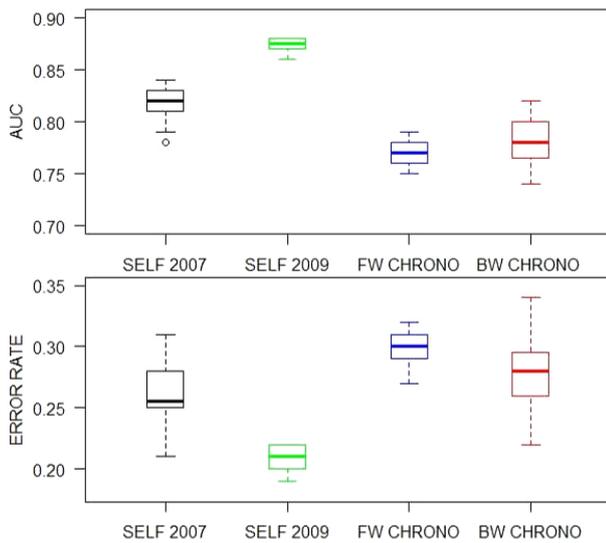
Figure 11. Distribution of the AUC and error rate values calculated on the 10 replicates for 2007 and 2009 modelling and 100 models during the chrono-validation process.
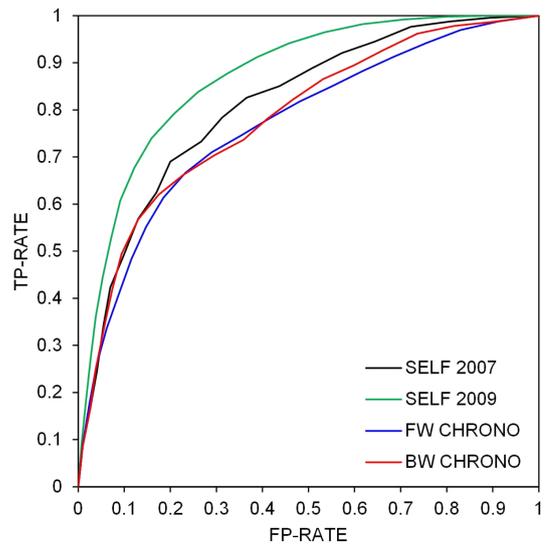
Figure 12. Comparison of the mean ROC curves obtained for the self-validated and chrono-validated models.

Table 3. HK values for the 100 replicated chrono-validations (in bold, the maximum values).

2007/2009				2009/2007			
Score	FP-rate	TP-rate	HK	Score	FP-rate	TP-rate	HK
0.990	0.000	0.000	0.000	0.970	0.000	0.000	0.000
0.941	0.010	0.089	0.080	0.925	0.010	0.086	0.076
0.898	0.024	0.175	0.151	0.890	0.026	0.166	0.139
0.862	0.040	0.259	0.219	0.860	0.038	0.250	0.212
0.815	0.062	0.337	0.275	0.820	0.057	0.340	0.283
0.764	0.089	0.411	0.322	0.778	0.074	0.419	0.345
0.723	0.116	0.484	0.368	0.729	0.094	0.495	0.400
0.681	0.147	0.552	0.404	0.646	0.131	0.568	0.438
0.635	0.185	0.614	0.429	0.558	0.174	0.620	**0.446**
0.581	0.233	0.666	**0.433**	0.481	0.228	0.662	0.434
0.518	0.290	0.710	0.420	0.394	0.296	0.704	0.408
0.457	0.354	0.746	0.392	0.323	0.359	0.737	0.378
0.402	0.416	0.783	0.366	0.265	0.410	0.781	0.371
0.350	0.482	0.818	0.336	0.219	0.466	0.822	0.356
0.304	0.549	0.850	0.300	0.176	0.532	0.865	0.334
0.263	0.617	0.882	0.266	0.143	0.598	0.895	0.296
0.224	0.686	0.913	0.227	0.113	0.662	0.927	0.265
0.183	0.757	0.942	0.186	0.081	0.737	0.962	0.225
0.138	0.829	0.970	0.140	0.055	0.816	0.978	0.162
0.087	0.912	0.987	0.076	0.030	0.903	0.987	0.084
0.014	1.000	1.000	0.000	0.007	1.000	1.000	0.000
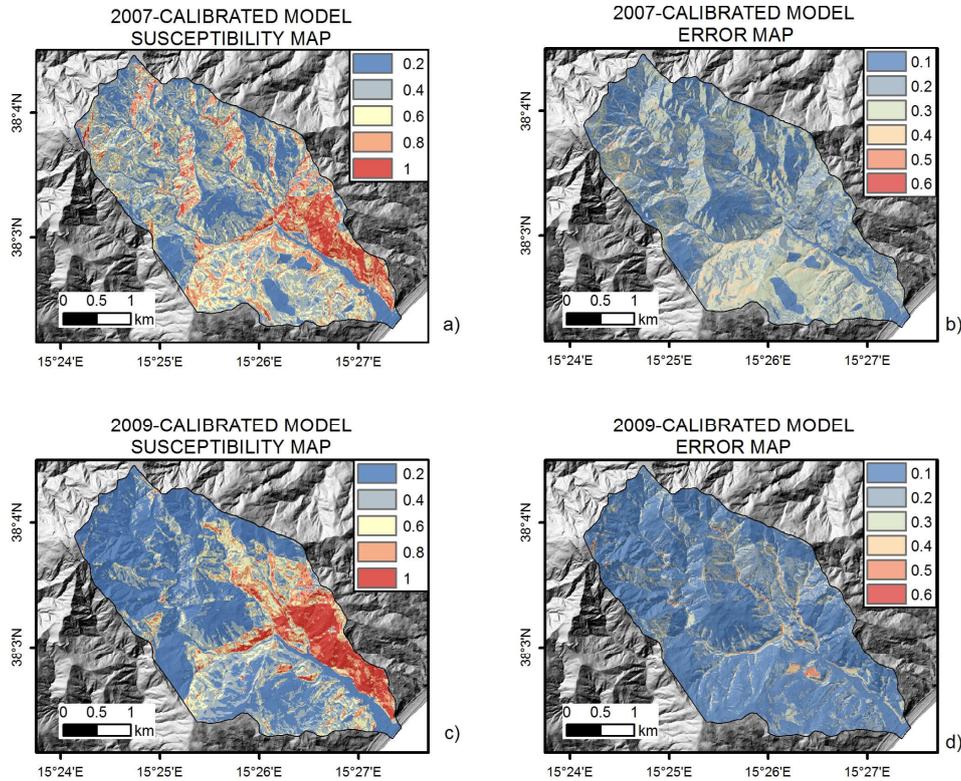
Figure 13. Susceptibility and error maps for the 2007 and the 2009 calibrated models: **(a)**, **(c)** mean susceptibility; **(b)**, **(d)** error maps.

Nat. Hazards Earth Syst. Sci., 15, 1785–1806, 2015

www.nat-hazards-earth-syst-sci.net/15/1785/2015/
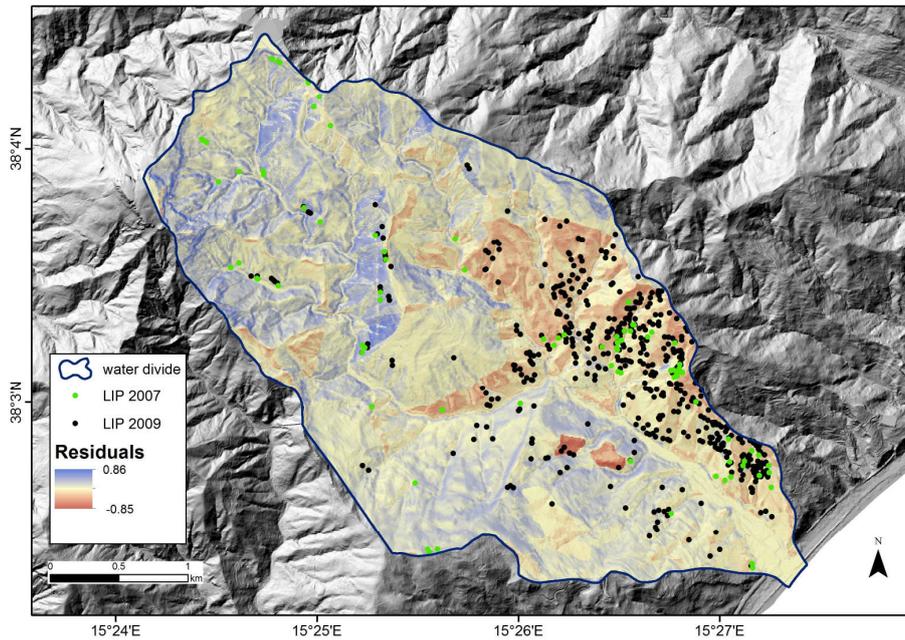
Figure 14. Map of residuals calculated as percentage differences between the two (2007 and 2009) mean susceptibilities.
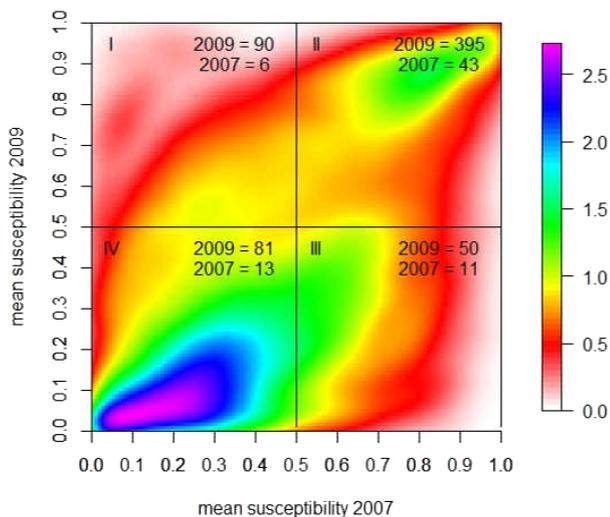
Figure 15. Dispersion density plot calculated using a 2-D binned kernel density algorithm (range for density calculation 0.045 xy). Positive cases for 0.5 cut-off values are reported for the two inventory events.

formances, a decrease in AUC from 0.81 to 0.77 for 2007, and from 0.87 to 0.78 for 2009, arose. Also, the mean error rate values increased from 0.26 to 0.30 for 2007, and from 0.20 to 0.28 for 2009. It is worth noting the strong decrease in performance affecting the 2009 model, which led the two chrono-validations to be almost equivalent. In Fig. 12, the calculated mean (over 100 replicates) ROC curves are shown. Coherently, the HK mean scores are comparable between for-

ward and backward validations, presenting a maximum of 0.433 and 0.446, respectively (Table 3).

A spatial view of the obtained prediction images for the 2007 and 2009 models is given in Fig. 13. In particular, the susceptibility maps show the spatial distribution of the mean probabilities for the 10 replicates, whilst the error maps describe the dispersion of the estimates, represented by a 2σ interval.

At a first glance, the two susceptibility maps appear quite different: the 2007 map shows a more diffused and graduated susceptibility, with the north-western and south-eastern sectors of the catchment hosting high susceptible areas. On the contrary, the 2009 map is characterized by a marked spatial separation between the north-eastern high susceptible sector and the remaining larger part of the catchment, which has a low susceptibility. In terms of error maps, the 2007 model is affected by a generally higher level of error, with the maximum values located in the central sector and minimum values along the stream network. The 2009 model, on the contrary, produced lower errors, with the exception of the stream network, which is characterized by relatively higher values, and two single small areas, corresponding to the outcrops of poorly diffused lithologies (see Fig. 1).

To compare the two landslide susceptibility maps, taking into consideration the distribution of the debris flows which occurred in 2007 and 2009, LIPs were located onto a map of the residuals. This map represents the difference between the two (2007 and 2009) mean susceptibilities (Fig. 14). The residuals confirmed the dissimilarity between the two models in estimating the susceptibility of the catchment, with higher probabilities in the southern and north-western sectors for the

forward-validated models, and in the north-eastern sector for the backward-validated models, respectively.

By comparing the two susceptibility estimates in a dispersion density plot (Fig. 15), the above-described trend is verified. The two models linearly agreed in the higher range of susceptibility, whilst a larger dispersion existed in the lower and intermediate susceptibility range. In particular, for the stable areas (near the origin of the plot) the higher densities pixels are shifted toward a more than 45° steep linear trend, marking an overestimation for the 2007 calibrated model.

From a binarized perspective, by setting the cut-off value for stable/unstable discrimination to 0.5, the final number of joint predictions (II, for TP, and IV, for FN sectors) was 77 %, whilst disjoint predictions (I and III sectors of the plot) reached 23 %. The two chrono-validated models performed in predicting the whole set of observed positives with different results: the backward-calibrated model produced $46 + 3$ (67 %) true positives and $13 + 11$ (33 %) false negatives for the 2007 LIPs, while the forward-calibrated model produced $395 + 50$ (72 %) true positives and $90 + 81$ (28 %) false negatives for the 2009 LIPs.

6 Discussion

In this study, the findings of previous studies (Zêzere et al., 2004; Guzzetti et al., 2005; Vergari et al., 2011; Wang et al., 2013) regarding the effectiveness of temporal partition procedures to explain future landslides are generally confirmed here, even in the case of debris flows triggered by an extreme rainfall event. The above-described results attest to a symmetry between forward and backward chrono-validations, as well as the main assumption on which stochastic modelling is based. However, through the analysis of the self-validated models, it was identified that the 2009 model resulted in a higher predictive performance, with a higher number of selected variables. This could be interpreted as a direct consequence of the greater number of debris flows which compose the 2009 inventory (1 order of magnitude more), so that a larger spectrum of multivariate conditions of the slopes was involved in failures and included in the data sets for the fitting of the models. However, the first four selected predictors for the 2009 model correspond to those composing the structure of the 2007 model: slope morphology (steepness, curvature and aspect), soil use and outcropping lithology.

The comparison between the performances of the self-validated and the chrono-validated models has highlighted a loss in accuracy which is slightly more marked for the higher performing self-validated 2009 model. Therefore, although a large difference between the accuracy of the two self-validated models is observed, the comparison between the forward and backward chrono-validated models shows very smoothed differences in terms of ROC-AUC and error rates. This suggests that, in spite of the higher performance which the 2009 model obtained in classifying the same 2009 event,

its skill in back-predicting the 2007 debris flow source areas is the same shown by the 2007 event in forward-predicting the debris flow source area of 2009.

The loss in performance demonstrated by the 2009 model suggests that using self-validated models for temporal prediction can mislead the user in estimating the performance of the model. In fact, one would expect that the model calibrated with the largest landslide inventory would be the best-performing in chrono-validation as it also includes the less extreme morphodynamic responses. However, in spite of the similar inner structure of the 2007 and 2009 models, the predictive performance of the 2009 backward model lowered to the same ROC-AUC and error rates of the 2007 forward model. The reason for this behaviour could be connected to the different local characteristics of the two storm events, which hit the slopes differently, even in such a small catchment. This would indicate, for this study case, that inside a 10 km^2 area there are two different pasts and two different futures, depending on which of the two storm events are used for calibration. This is a similar finding to that obtained for chrono-validation procedures by Chang et al. (2014) in a larger-scale (tropical cyclones) study, whose predictive models even resulted in being "capable of predicting landslides triggered by a strong typhoon but not a weak typhoon" (i.e. their best model missed nearly all the landslide cells triggered by the weak typhoon events).

At the same time, a non-linearity of the morphodynamic response of the slopes (different coefficients and/or predictors) could affect the performance in chrono-validation: a larger event does not produce a larger response which include less intense storms, but rather a different one. The larger the difference between the triggering events, the greater the distinction in the response of the preparatory conditions.

In the domain of the predictors, this is highlighted by the different inner structures of the models. If compared to the 2007 event, the 2009 event also activated eastern and south-eastern-facing pixels, as well as high metamorphic-grade (MLEa) lithologies and terraced deposits; topographic curvatures, distance from faults and soil use (the latter with negative coefficients) have also taken an important role in controlling the distribution of the debris flow source areas. However, this richer structure of the model does not increase its predictive ability with respect to the distribution of the 2007 debris flows; the backward chrono-validation does not demonstrate this greater accuracy. This suggests that the 2007 debris flows were activated through different, even if largely overlapping, mechanisms.

In the domain of the geographical space, the map of the residuals provided a spatial view of the different behaviour of the two models, giving the interpreter clues for the real path followed by the two storm fronts inside the Itala catchment. The 2009 model markedly overestimated the susceptibilities in the central-northern sector of the catchment, whilst the 2007 model produced higher susceptibilities than 2009 in the north-western inner mountain sector. Regardless of the

different intensities, this spatial trend suggests that the 2009 storm path was limited to the coastal area, whilst the 2007 storm affected the whole catchment more homogeneously, activating also the slopes of the mountain sector. This interpretation is also confirmed by the different spatial distribution of the debris flows of the two event inventories and it agrees with the findings on a much larger scale of Chang et al. (2014).

However, from a risk perspective, the difference between the two models did not produce a significant loss in prediction, as only a limited number of cases resulted in a false positive prediction. This is why the mapped debris flows are largely located in the more susceptible pixels. However, the results of the present research have confirmed that the larger difference between the two models has been observed in the intermediate susceptibilities interval, which is the same region of the error plots where the self-validated models show poor precision. This difference is also attested by the HK scores, which confirmed the good prediction skills, but with maximum values proximal to 0.45. Under the considered triggering conditions, the multivariate relationship between debris flow activation and predictors is in fact linear, so that no single marked cut-off value for probability accurately discriminates positives from negatives. Nevertheless, it is worth highlighting the selection of a 0.5 cut-off value, which resulted in a higher performance for the temporal prediction of the positive cases (forward chrono-validation) of the 2007 calibrated model.

Finally, it is worth comparing here the results obtained for chrono-validation (AUC was 0.77 / 0.78), with the ones from Lombardo et al. (2014), which applied a spatial exportation scheme in two catchments very close to each other. In fact, a higher performance (AUC was 0.83) resulted for the prediction skill of the transferability procedure which was adopted there, by calibrating the model in the Briga catchment to predict the Giampilieri debris flows, using event inventories produced by the same 2009 storm-triggering event. Sharing the triggering event allows for a better predictive ability, in spite of the circumstance that, in a spatial partition scheme, the calibrated model is totally blind with respect to the validation area, in terms of the spatial combination of the predictors and the target pattern (the unknown debris flows).

7 Conclusions

The results obtained in this research confirmed that the basic assumption on which susceptibility modelling is based (the past is the key to the future) must be critically accepted in the case of extreme events. In fact, in the case of the two storm events considered here, the dissimilarities in the intensity and the real path of the two storm fronts produced measurable differences in the behaviour of the two derived predictive models, both in the domain of the predictors and in the spatial pattern of the susceptibility maps. Two main causes

have been recognized here: on the one hand, the slopes did not linearly respond to the trigger intensity, so that different predictors and coefficients were fitted by the two regressed models; on the other hand, effects produced by the spatial non-homogeneity of the rain intensity for each single storm event, even at the scale of such small catchments, were detected.

In terms of the operative use of the susceptibility maps, the effects identified attest to the risk of either over- or underestimating the susceptibility, both for the 2009 and 2007 models. In particular, limits arise in the general perspective of using the most severe and available inventory for calibrating the best-performing model. In fact, in this research it was verified that this best-performing self-validated model did not result in the most accurate one in chrono-validation, also demonstrating susceptibility underestimation and false negative production.

In the present study, the differences between the two models basically reside in the intermediate susceptibility interval, so that a precautionary approach in reclassifying the susceptibility map could be adopted, accepting the precision limits in the intermediate probability classes. However, larger differences between the triggering storms to which calibration and validation event inventories are connected could result in larger predictive limits and more misleading susceptibility maps.

The strict relation between trigger intensity, slope response and prediction performance arises also from the comparison of this study to another study carried out by applying spatial partition or transferability validation strategies in two adjacent catchments for the same 2009 trigger, obtaining a better predictive performance. In the opinion of the authors, this difference confirms limitations of the chrono-validation procedure when working with extreme rainfall events. For this reason, the application of transferability or chrono-validation should be evaluated from time to time on the basis of the availability of historical records of phenomena, information on the trigger event, and similarity with other areas where debris flow events have already occurred. At the same time, the production of susceptibility maps such as those presented in this paper constitutes a basic starting point for modelling propagation, run-out and magnitude associated to the predicted phenomena, so that an estimation of the debris flow hazard is achieved within a given area.

References

Ardizzone, F., Basile, G., Cardinali, M., Casagli, N., Del Conte, S., Del Ventisette, C., Fiorucci, F., Garfagnoli, F., Gigli, G., Guzzetti, F., Iovine, G., Mondini, A. C., Moretti, S., Panebianco, M., Raspini, F., Reichenbach, P., Rossi, M., Tanteri, L., and Terranova, O.: Landslide inventory map for the Briga and the Giampilieri catchments, NE Sicily, Italy, J. Maps, 8, 176–180, doi:10.1080/17445647.2012.694271, 2012.

Aronica, G. T., Brigandí, G., and Morey, N.: Flash floods and debris flow in the city area of Messina, north-east part of Sicily, Italy in October 2009: The case of the Giampilieri catchment, Nat. Hazards Earth Syst. Sci., 12, 1295–1309, doi:10.5194/nhess-12-1295-2012, 2012a.

Aronica, G. T., Biondi, G., Brigandì, G., Cascone, E., Lanza, S., and Randazzo, G.: Assessment and mapping of debris-flow risk in a small catchment in eastern Sicily through integrated numerical simulations and GIS, Phys. Chem. Earth, Parts A/B/C, 49, 52–63, doi:10.1016/j.pce.2012.04.002, 2012b.

Atkinson, P., Jiskoot, H., Massari, R., and Murray, T.: Generalised linear modelling in geomorphology, Earth Surf. Process. Landforms, 23, 1185–1195, 1998.

Atkinson, P. M. and Massari, R.: Autologistic modelling of susceptibility to landsliding in the Central Apennines, Italy, Geomorphology, 130, 55–64, doi:10.1016/j.geomorph.2011.02.001, 2011.

Bai, S. B., Wang, J., Lü, G. N., Zhou, P. G., Hou, S. S., and Xu, S. N.: GIS-based and data-driven bivariate landslide-susceptibility mapping in the Three Gorges Area, China, Pedosphere, 19, 14–20, doi:10.1016/S1002-0160(08)60079-X, 2009.

Beguería, S.: Changes in land cover and shallow landslide activity: A case study in the Spanish Pyrenees, Geomorphology, 74, 196–206, doi:10.1016/j.geomorph.2005.07.018, 2006.

Bordoni, M., Meisina, C., Valentino, R., Bittelli, M., and Chersich, S.: Site-specific to local-scale shallow landslides triggering zones assessment using TRIGRS, Nat. Hazards Earth Syst. Sci., 15, 1025–1050, doi:10.5194/nhess-15-1025-2015, 2015.

Brabb, E. E.: Innovative approaches to landslide hazard and risk mapping, Proceedings of the 4th International Symposium on Landslides, 16–21 September, Toronto, Ontario, Canada (Canadian Geotechnical Society, Toronto, Ontario, Canada), 1, 307–324, 1984.

Brenning, A.: Spatial prediction models for landslide hazards: review, comparison and evaluation, Nat. Hazards Earth Syst. Sci., 5, 853–862, doi:10.5194/nhess-5-853-2005, 2005.

Carrara, A., Cardinali, M., Guzzetti, F., and Reichenbach, P.: GIS technology in mapping landslide hazard, edited by: Carrara, A., and Guzzetti, F., in: Geographical Information Systems in Assessing Natural Hazards, Kluwer Academic Publisher, Dordrecht, The Netherlands, 135–175, 1995.

Carrara, A., Crosta, G., and Frattini, P.: Geomorphological and historical data in assessing landslide hazard, Earth Surf. Process. Landforms, 28, 1125–1142, doi:10.1002/esp.545, 2003.

Carrara, A., Crosta, G., and Frattini, P.: Comparing models of debris-flow susceptibility in the alpine environment, Geomorphology, 94, 353–378, doi:10.1016/j.geomorph.2006.10.033, 2008.

Chang, K., Chiang, S., Chen, Y., and Mondini, C.: Modeling the spatial occurrence of shallow landslides triggered by typhoons, Geomorphology, 137–148, doi:10.1016/j.geomorph.2013.11.020, 2014.

Chung, C.-J. F. and Fabbri, A. G.: Validation of Spatial Prediction Models for Landslide Hazard Mapping, Nat. Hazards, 30, 451–472, doi:10.1023/B:NHAZ.0000007172.62651.2b, 2003.

Ciampalini, A., Raspini, F., Bianchini, S., Frodella, W., Bardi, F., Lagomarsino, D., Di Traglia, F., Moretti, S., Proietti, C., Pagliara, P., Onori, R., Corazza, A., Duro, A., Basile, G., and Casagli, N.: Remote sensing as tool for development of landslide databases: The case of the Messina Province (Italy) geodatabase, Geomorphology, 1–16, doi:10.1016/j.geomorph.2015.01.029, 2015.

Conoscenti, C., Di Maggio, C., and Rotigliano, E.: GIS analysis to assess landslide susceptibility in a fluvial basin of NW Sicily (Italy), Geomorphology, 94, 325–339, doi:10.1016/j.geomorph.2006.10.039, 2008a.

Conoscenti, C., Maggio, C., and Rotigliano, E.: Soil erosion susceptibility assessment and validation using a geostatistical multivariate approach: A test in Southern Sicily, Nat. Hazards, 46, 287–305, doi:10.1007/s11069-007-9188-0, 2008b.

Conrad, O.: SAGA – Entwurf, Funktionsumfang und Anwendung eines Systems für Automatisierte Geowissenschaftliche Analysen, Dissertation, University of Göttingen, 221 pp., 2007.

Costanzo, D., Cappadonia, C., Conoscenti, C., and Rotigliano, E.: Exporting a Google Earth™ aided earth-flow susceptibility model: A test in central Sicily, Nat. Hazards, 61, 103–114, doi:10.1007/s11069-011-9870-0, 2012a.

Costanzo, D., Rotigliano, E., Irigaray, C., Jiménez-Perálvarez, J. D., and Chacón, J.: Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: Application to the river Beiro basin (Spain), Nat. Hazards Earth Syst. Sci., 12, 327–340, doi:10.5194/nhess-12-327-2012, 2012b.

Costanzo, D., Chacón, J., Conoscenti, C., Irigaray, C., and Rotigliano, E.: Forward logistic regression for earth-flow landslide susceptibility assessment in the Platani river basin (southern Sicily, Italy), Landslides, 11, 639–653, doi:10.1007/s10346-013-0415-3, 2014.

De Guidi, G. and Scudero, S.: Landslide susceptibility assessment in the Peloritani Mts. (Sicily, Italy) and clues for tectonic control of relief processes, Nat. Hazards Earth Syst. Sci., 13, 949–963, doi:10.5194/nhess-13-949-2013, 2013.

Del Ventisette, C., Garfagnoli, F., Ciampalini, A., Battistini, A., Gigli, G., Moretti, S., and Casagli, N.: An integrated approach to the study of catastrophic debris-flows: Geological hazard and human influence, Nat. Hazards Earth Syst. Sci., 12, 2907–2922, doi:10.5194/nhess-12-2907-2012, 2012.

Frattini, P., Crosta, G., and Carrara, A.: Techniques for evaluating the performance of landslide susceptibility models, Eng. Geol., 111, 62–72, doi:10.1016/j.enggeo.2009.12.004, 2010.

Gariano, S. L., Brunetti, M. T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O., Vennari, C., and Guzzetti, F.: Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy, Geomorphology, 228, 653–665, doi:10.1016/j.geomorph.2014.10.019, 2015.

Goswami, R., Mitchell, N. C., and Brocklehurst, S. H.: Distribution and causes of landslides in the eastern Peloritani of NE Sicily and western Aspromonte of SW Calabria, Italy, Geomorphology, 132, 111–122, doi:10.1016/j.geomorph.2011.04.036, 2011.

Guzzetti, F.: Landslide Hazard and Risk Assessment, PhD dissertation, Rheinischen Friedrich-Wilhelms-Universität Bonn, Germany, p. 373, available at: http://hss.ulb.uni-bonn.de/2006/0817/0817.html (last access: 7 August 2015), 2005.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., and Ardizzone, F.: Probabilistic landslide hazard assessment at the basin scale, Geomorphology, 72, 272–299, doi:10.1016/j.geomorph.2005.06.002, 2005.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., and Galli, M.: Estimating the quality of landslide susceptibility models, Geomorphology, 81, 166–184, doi:10.1016/j.geomorph.2006.04.007, 2006.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K. T.: Landslide inventory maps: New tools for an old problem, Earth-Sci. Rev., 112, 42–66, 2012.

Hanssen, A. W. and Kuipers, W. J. A.: On the relationship between the frequency of rain and various meteorological parameters, Meded. Verh., 81, 2–15, 1965.

Heckmann, T., Gegg, K., Gegg, A., and Becht, M.: Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows, Nat. Hazards Earth Syst. Sci., 14, 259–278, doi:10.5194/nhess-14-259-2014, 2014.

Hosmer, D. W. and Lemeshow, S.: Applied logistic regression, Wiley Ser. Probab. Stat. 2nd, 375, doi:10.1198/tech.2002.s650, 2000.

Hungr, O., Evans, S. G., Bovis, M. J., and Hutchinson, J. N.: A review of the classification of landslides of the flow type, Environ. Eng. Geosci., 7, 221–238, doi:10.2113/gseegeosci.7.3.221, 2001.

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

Hutchinson, J. N.: General Report: Morphological and geotechnical parameters of landslides in relation to geology and hydrogeology, Proceedings, Fifth International Symposium on Landslides, edited by: Bonnard, C., 1, 3–35, Rotterdam, Balkema, 1988.

Köppen, W.: Die Klimate der Erde, Walter de Gruyter, Berlin, 1–369, 1923.

Lagomarsino, D., Segoni, S., Rosi, A., Rossi, G., Battistini, A., Catani, F., and Casagli, N.: Quantitative comparison between two different methodologies to define rainfall thresholds for landslide forecasting, Nat. Hazards Earth Syst. Sci. Discuss., 3, 891–917, doi:10.5194/nhessd-3-891-2015, 2015.

Lentini, F., Carbone, S., and Messina, A.: Carta Geologica d'Italia scala 1:50.000 Foglio 601 "Messina-Reggio di Calabria", con note illustrative. Apat Agenzia per la protezione dell'ambiente e per i servizi tecnici, Dipartimento Difesa del Suolo – Servizio geologico d'Italia, 2007.

Lombardo, L., Cama, M., Maerker, M., and Rotigliano, E.: A test of transferability for landslides susceptibility models under extreme climatic events: application to the Messina 2009 disaster, Nat. Hazards, 74, 1951–1989, doi:10.1007/s11069-014-1285-2, 2014.

Menard, S.: Applied Logistic Regression Analysis, SAGE, 1–120, doi:10.4135/9781412983433, 2002.

Meusburger, K. and Alewell, C.: On the influence of temporal change on the validity of landslide susceptibility maps, Nat. Hazards Earth Syst. Sci., 9, 1495–1507, doi:10.5194/nhess-9-1495-2009, 2009.

Messina, A., Somma, R., Macaione, E., Carbone, G., and Careri, G.: Peloritani continental crust composition (southern Italy): geological and petrochemical evidences, Boll. Soc. Geol. Ital., 123, 405–444, 2004.

Mondini, a. C., Guzzetti, F., Reichenbach, P., Rossi, M., Cardinali, M., and Ardizzone, F.: Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images, Remote Sens. Environ., 115, 1743–1757, doi:10.1016/j.rse.2011.03.006, 2011.

Moore, I. D., Gessler, P. E., Nielsen, G. A., and Peterson, G. A.: Soil Attribute Prediction Using Terrain Analysis, Soil Sci. Soc. Am. J., 57, NP, doi:10.2136/sssaj1993.572NPb, 1993.

Nefeslioglu, H. A., Gokceoglu, C., and Sonmez, H.: An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps, Eng. Geol., 97, 171–191, doi:10.1016/j.enggeo.2008.01.004, 2008.

Ohlmacher, G. C. and Davis, J. C.: Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA, Eng. Geol., 69, 331–343, doi:10.1016/S0013-7952(03)00069-3, 2003.

Ohlmacher, G. C.: Plan curvature and landslide probability in regions dominated by earth flows and earth slides, Eng. Geol., 91, 117–134, doi:10.1016/j.enggeo.2007.01.005, 2007.

Othman, A. A., Gloaguen, R., Andreani, L., and Rahnama, M.: Landslide susceptibility mapping in Mawat area, Kurdistan Region, NE Iraq: a comparison of different statistical models, Nat. Hazards Earth Syst. Sci. Discuss., 3, 1789–1833, doi:10.5194/nhessd-3-1789-2015, 2015.

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

Petschko, H., Brenning, A., Bell, R., Goetz, J., and Glade, T.: Assessing the quality of landslide susceptibility maps – case study Lower Austria, Nat. Hazards Earth Syst. Sci., 14, 95–118, doi:10.5194/nhess-14-95-2014, 2014.

Rakotomalala, R.: Tanagra: un logiciel gratuit pour l'enseignement et la recherche, in: Actes De EGC, 697–702, 2005.

Reichenbach, P., Busca, C., Mondini, A. C., and Rossi, M.: The influence of land use change on landslide susceptibility zonation: the Briga Catchment Test Site (Messina, Italy), Environ. Manage., 54, 1372–1384, doi:10.1007/s00267-014-0357-0, 2014.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A. C., and Peruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecasts, Geomorphology, 114, 129–142, doi:10.1016/j.geomorph.2009.06.020, 2010.

Rotigliano, E., Agnesi, V., Cappadonia, C., and Conoscenti, C.: The role of the diagnostic areas in the assessment of landslide susceptibility models: A test in the sicilian chain, Nat. Hazards, 58, 981–999, doi:10.1007/s11069-010-9708-1, 2011.

Schraml, K., Thomschitz, B., McArdell, B. W., Graf, C., and Kaitna, R.: Modeling debris-flow runout patterns on two alpine fans with different dynamic simulation models, Nat. Hazards Earth Syst. Sci., 15, 1483–1492, doi:10.5194/nhess-15-1483-2015, 2015.

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

Stancanelli, L. M. and Foti, E.: A comparative assessment of two different debris flow propagation approaches – blind simulations on a real debris flow event, Nat. Hazards Earth Syst. Sci., 15, 735–746, doi:10.5194/nhess-15-735-2015, 2015.

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

Süzen, M. L. and Doyuran, V.: A comparison of the GIS based landslide susceptibility assessment methods: Multivariate versus bivariate, Environ. Geol., 45, 665–679, doi:10.1007/s00254-003-0917-8, 2004.

Tseng, C. M., Lin, C. W., and Hsieh, W. D.: Landslide susceptibility analysis by means of event-based multi-temporal landslide inventories, Nat. Hazards Earth Syst. Sci. Discuss., 3, 1137–1173, doi:10.5194/nhessd-3-1137-2015, 2015.

Van Den Eeckhaut, M., Reichenbach, P., Guzzetti, F., Rossi, M., and Poesen, J.: Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium, Nat. Hazards Earth Syst. Sci., 9, 507–521, doi:10.5194/nhess-9-507-2009, 2009.

Varnes, D. J.: Slope movement: types and process, edited by: Schuster, R. L. and Krizek, R. J., Landslides: Analyses and control, Transportation Research Board, Natural Academy of Science, Washington, Sp. Rep., 176, 11–33, 1978.

Vergari, F., Della Seta, M., Del Monte, M., Fredi, P., and Lupia Palmieri, E.: Landslide susceptibility assessment in the Upper Orcia Valley (Southern Tuscany, Italy) through conditional analysis: A contribution to the unbiased selection of causal factors, Nat. Hazards Earth Syst. Sci., 11, 1475–1497, doi:10.5194/nhess-11-1475-2011, 2011.

Von Ruette, J., Papritz, A., Lehmann, P., Rickli, C., and Or, D.: Spatial statistical modeling of shallow landslides-Validating predictions for different landslide inventories and rainfall events, Geomorphology, 133, 11–22, 2011.

Wang, X., Zhang, L., Wang, S., and Lari, S.: Regional landslide susceptibility zoning with considering the aggregation of landslide points and the weights of factors, Landslides, 11, 399–409, doi:10.1007/s10346-013-0392-6, 2014.

Zêzere, J. L., Reis, E., Garcia, R., Oliveira, S., Rodrigues, M. L., Vieira, G., and Ferreira, A. B.: Integration of spatial and temporal data for the definition of different landslide hazard scenarios in the area north of Lisbon (Portugal), Nat. Hazards Earth Syst. Sci., 4, 133–146, doi:10.5194/nhess-4-133-2004, 2004.