

Handling high predictor dimensionality in slope-unit-based landslide susceptibility models through LASSO-penalized Generalized Linear Model

Daniela Castro Camilo ^a, Luigi Lombardo ^{a,b,c,*}, P. Martin Mai ^b, Jie Dou ^d, Raphaël Huser ^a

^a CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Jeddah, Saudi Arabia

^b PSE Division, King Abdullah University of Science and Technology, Thuwal, Jeddah, Saudi Arabia

^c Department of Geosciences, Tübingen University, Tübingen, Germany

^d Public Works Research Institute, Japan



ARTICLE INFO

Article history:

Received 20 December 2016

Received in revised form

5 July 2017

Accepted 8 August 2017

Keywords:

High predictor dimensionality

Slope Unit-based landslide susceptibility

Lasso penalized variable selection

R scripting

Open sourcing

ABSTRACT

Grid-based landslide susceptibility models at regional scales are computationally demanding when using a fine grid resolution. Conversely, Slope-Unit (SU) based susceptibility models allows to investigate the same areas offering two main advantages: 1) a smaller computational burden and 2) a more geomorphologically-oriented interpretation. In this contribution, we generate SU-based landslide susceptibility for the Sado Island in Japan. This island is characterized by deep-seated landslides which we assume can only limitedly be explained by the first two statistical moments (mean and variance) of a set of predictors within each slope unit. As a consequence, in a nested experiment, we first analyse the distributions of a set of continuous predictors within each slope unit computing the standard deviation and quantiles from 0.05 to 0.95 with a step of 0.05. These are then used as predictors for landslide susceptibility. In addition, we combine shape indices for polygon features and the normalized extent of each class belonging to the outcropping lithology in a given SU. This procedure significantly enlarges the size of the predictor hyperspace, thus producing a high level of slope-unit characterization. In a second step, we adopt a LASSO-penalized Generalized Linear Model to shrink back the predictor set to a sensible and interpretable number, carrying only the most significant covariates in the models. As a result, we are able to document the geomorphic features (e.g., 95% quantile of Elevation and 5% quantile of Plan Curvature) that primarily control the SU-based susceptibility within the test area while producing high predictive performances. The implementation of the statistical analyses are included in a parallelized R script (LUDARA) which is here made available for the community to replicate analogous experiments.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

The landslide susceptibility (LS, hereafter) represents the relative spatial probability of landslide occurrence (Chung and Fabbri, 1999; Guzzetti et al., 2006b) that is typically computed through different statistical approaches and over a geographic space which can be discretized into different mapping units (Carrara et al., 1995; Guzzetti et al., 1999).

A mapping unit refers to a portion of land surface with analogous geologic and/or geomorphic properties that differ from its surroundings across definable boundaries (Hansen, 1984). The

advancements in LS assessment over the last three decades have primarily reflected developments on the two aforementioned topics where: 1) various statistical approaches have been tested (Cama et al., 2015; Conoscenti et al., 2016), compared (Felicísimo et al., 2013; Süzen and Doyuran, 2004) and reviewed (Aleotti and Chowdhury, 1999; Brenning, 2005; Van Westen et al., 2006) while 2) different mapping units have been adopted (Galli et al., 2008; Van Westen, 2000), their respective effects evaluated (Arnone et al., 2016; Carrara et al., 2008; Van Den Eeckhaut et al., 2009) and broadly summarized into four categories: grid cells (Lombardo et al., 2015), terrain units (Calvello et al., 2013), unique

* Corresponding author. CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Jeddah, Saudi Arabia.

E-mail address: luigi.lombardo83@gmail.com (L. Lombardo).

condition units (Chung et al., 1995) and slope units (SU) (Guzzetti and Reichenbach, 1994). Each of these mapping units offer advantages and drawbacks which were already described by Carrara et al. (1995). Landslide automatic recognition (McKean and Roering, 2004; Nichol and Wong, 2005; Tarolli et al., 2012), remote sensing (Abdulwahid and Pradhan, 2016; Lombardo et al., 2016a) and deterministic (An et al., 2016; Iovine et al., 2007) based predictions are also regarded as important topics within the geomorphological community for susceptibility purposes.

Despite the long-term efforts and the rich literature, the adoption of the best mapping unit when assessing the LS is still unclear (Guzzetti et al., 2000). Few examples exist where different mapping units were tested for the same area. Van Den Eeckhaut et al. (2009) compared grid-based and SU-based structures obtaining analogous spatial patterns and performances. In particular, the grid structure had a slightly higher prediction skill but a lower robustness. Conversely, Erener and Düzgün (2012) denoted an increase in predictive performances when using the SU. The reason behind these results may be suggested by Rotigliano et al. (2012) who highlighted that single cell values are less representative when considering phenomena involving portion or whole slopes, thus making the SU the correct spatial domain to operate upon. Assuming that SU is a reliable mean for partitioning a given area for landslide spatial prediction purposes, we conducted a literature review on the topic. We found that the whole community operates by summarizing the n -pixels contained in a given SU by using just their mean and variance or standard deviation (e.g., Guzzetti et al., 2006a; Rossi et al., 2010; Zhou et al., 2015) and in fewer cases adding the range (Reichenbach et al., 2014). The distribution of the adopted predictors within each SU has never been investigated more in detail, despite the fact that thousands of pixels may be contained in a SU, making their representation through just two values potentially inadequate.

Following this observation, we first hypothesize that the two aforementioned statistical moments within a given slope unit can only offer part of the information for modelling large landslides. To be able to capture potentially non-linear effects, and to better assess the influence of the *whole* predictors' distribution, we represent each continuous predictor by their 5th to 95th percentiles (computed for each SU) using a 5% quantile step. As a result of this procedure, the size of the predictor hyperspace proportionally increased by one order of magnitude and high multicollinearity among predictors was induced. In order to reduce and extract only the most relevant covariates, we secondly applied a LASSO (Least Absolute Shrinkage and Selection Operator) penalized Generalized Linear Model. This is also different from the other contributions in the literature as the community unanimously adopts a stepwise framework for variable selection (Baeza and Corominas, 2001; Cama et al., 2016; Capitani et al., 2013; Goetz et al., 2015; Süzen and Kaya, 2012). The advantage of LASSO is that model selection and parameter estimation are performed simultaneously, while the LASSO penalty can be adjusted for optimal prediction.

As a result, the main novelty in this paper consists of testing the general assumption that mean and variance are good indicators of the whole distribution of any continuous covariate in the model within a given slope unit. If this assumption holds we expect median or near-median quantiles to emerge from the variable selection process. However, if quantiles significantly different from the median are selected, then the final susceptibility might differ from the one generated with the traditional mean values. In this framework, the introduction of an untested variable selection tool also represents a further improvement in the literature as the stepwise selection procedure is already known for many weaknesses in the statistical community despite its wide use in the geomorphological one.

2. Materials and methods

2.1. Study area

The study area is located in a mountainous region of Sado Island of Niigata Prefecture, in the Japan Sea (Fig. 1 (a)). Sado Island is composed of two parallel, elongated ridges approximately in a northeast to southwest direction. These are locally named Osado and Kosado and cover approximately 856 km². The climate is humid, subtropical with warm summers and cold winters. The average temperature is 13.7 °C and the mean annual rainfall is around 1550 mm (Yamagishi, 2008). The elevation varies from sea level to 1172 m with a mean of 333 m. The highest point of the island coincides with the peak of Mt. Kimpoku in Osado. The local geology is primarily represented by Neogene terrestrial and marine volcanic rocks of rhyolitic, dacitic and andesitic nature associated with pyroclastites and intrusives rhyolites. Most of the coastal slopes are rocky, and some shores are recently formed by semi-consolidated and unconsolidated sand and gravel. Being characterized by mountains, hills upheaved benches and active faults, the study area is prone to landsliding and several damages are reported (Yamagishi, 2008). In the Osado area, the primary cause for landslides is represented by rainfall and partly by snow melting.

The exposure to high intensity precipitation is mainly concentrated between June–July and November–January as shown in Fig. 2 where TRMM rainfall estimates (<https://pmm.nasa.gov/trmm>) are summarized over 14 years to highlight seasonal trends. The aforementioned wet intervals coincide with times when displacements have been measured by Ayalew et al. (2005a) for specific landslide cases coinciding with strong groundwater level rise. In particular, Ayalew et al. (2005a) highlight the role of heavy rainfall between October 2000, and January 2001, for the mobilization of Shiidomari landslide located to the East of the Island.

2.2. Landslide inventory

In this contribution we test our method using a well-studied

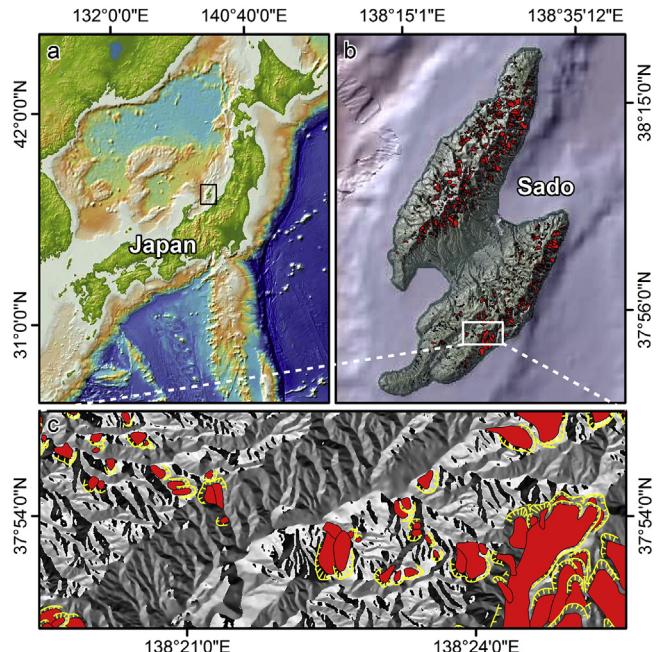


Fig. 1. Study area: Geographic location (a), Sado Island (b), detail of the landslide inventory (red polygons) and head scarps (yellow polylines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

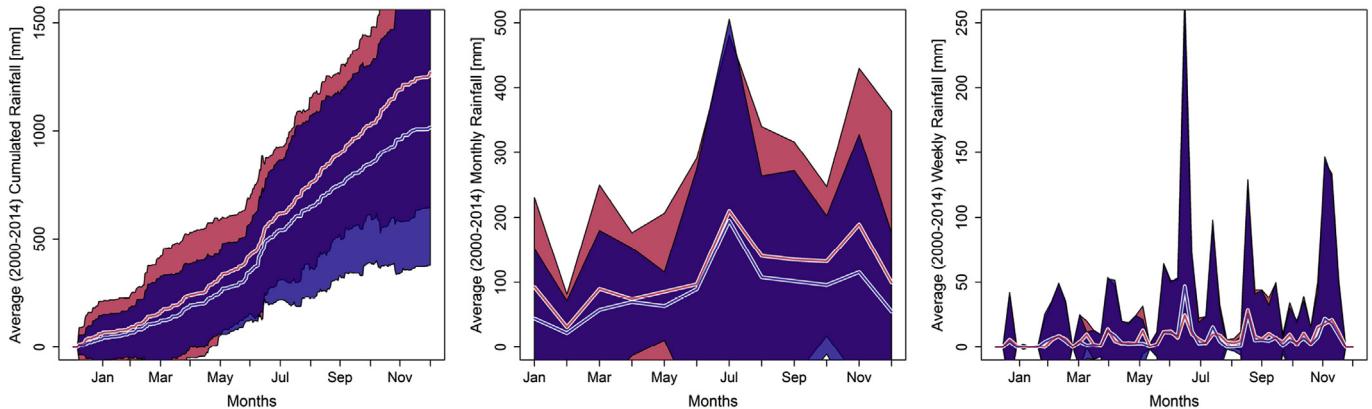


Fig. 2. Average rainfall cumulative curves computed between years 2000–2014 (a). Average monthly rainfall computed between years 2000–2014 (b). Average weekly rainfall computed between years 2000–2014 (c). Light blue colors refer to the TRMM pixel falling to the North of the Sado island whereas red represents TRMM data from the south. Lines represent mean data across 14 years while areas are 95% confidence intervals. Overlapping confidence intervals are shown in a purple color. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

landslide archive. The archive for Sado Island (Fig. 1(b)) provided such conditions as it has been validated over several articles (Ayalew et al., 2005a,b; Dou et al., 2015b,a). The landslide inventory was built by the National Research Institute for Earth Science and Disaster Prevention (NIED) in Japan, being mapped through multiple aerial photographs since the year 2000. The NIED digitized the landslides scars and made them available as polygon shapefiles (Fig. 1(c)). 1538 landslides are mapped within the island affecting both the ridges with analogous distributions along the rocky coastline whilst a greater number of occurrences characterizes the northern area. They account for a total surface of 91.5 km^2 which represents 11.1% of the whole surface area. From a geomorphometric perspective (Fig. 3(a)-(f)) the landslide distribution is positively skewed and heavy-tailed, with the majority of the mass movements having similar characteristics and fewer much more extreme cases due to their size. In term of individual sizes (Fig. 3(a)), 84% of the landslides extent for more than 10000 m^2 . This attests for the significant area exposed to failures and indicates the almost unique landslide class considered in the present study.

The landslide types are reported as deep-seated, either rotational or translational slides, and a limited number of the rock falls which were excluded for the present research. The geological control over landslides appears evident when intersecting the landslide scars with the underlying geology (Fig. 3(g)). Overall, 65% of the landslides occur on volcanic andesite followed by 23% on dacite and 7.6% on sandstone. These lithotypes are severely fractured as the whole island is subjected to active faulting since the early and middle Miocene when the extensional stress due to the Japan Sea opening (Jolivet et al., 1992) reached the Island producing the graben structure between the Osado and Kosado ridges. These lineaments cuts through Oligocene to Middle Miocene volcanics and volcanoclastic rocks (Ganzawa, 1983). Towards the end of the Miocene, the stress regime even switched to compressive in conjunction with the interruption of the sea opening (Jolivet et al., 1991) giving rise to thrusts that further weakened the rock mass. The whole inventory provided by NIED lacks systematic classification of landslide typologies as their sizes hinder investigations for each case. This is actually available for few landslides while the majority is defined as deep-seated without discriminating between rotational or translational mechanisms. Furthermore, the inventory lacks the temporal information necessary for evolution assessments in relation to the precipitation regime.

2.3. Model building strategy

2.3.1. Slope units and status assignment

Among the mapping units, a SU represents the region of space delimited between ridges and valleys under the constraint of homogeneous slope aspect and steepness distributions (Carrara et al., 1991; Tian et al., 2010). Thus, it physically describes either the left or right side of a sub-basin of any order into which a watershed is subdivided (Erener and Düzgün, 2012). As a consequence, it naturally reflects the morphodynamic response of a given slope in case of landslide occurrence (Huabin et al., 2005) making it possible to avoid the shortcomings of low geomorphological representativeness of grid-based susceptibility mapping. This property makes the SU particularly suitable for modelling slope instabilities due to large landslides (Beguería and Lorente, 2007). Therefore, we adopt the SU as a mean to represent the deep-seated instabilities dominating the failure mechanisms within the Sado Island. It is worth mentioning that SU has been commonly used for rainfall triggered landslides whereas few contributions can be found on their seismically induced counterpart (e.g., Zhou et al., 2015). The reason for this may be due to the actual size of the latter type spanning over more than one slope unit, thus adding complexity to the mapping unit choice.

The SUs (Fig. 4) have been computed by using the *r.slopeunits* software developed at the Geomorphology Research Group (<http://geomorphology.ipi.cnr.it/tools/slope-units>) and recently explained in details by Alvioli et al. (2016). This Python code is accessible through GRASS GIS (Neteler and Mitasova, 2013) and allows for automatic delineation of SUs from a Digital Elevation Model (DEM). The code initially uses the *r.watershed* script (Metz et al., 2011) to obtain a first-order delineation of half-basins. Subsequently, it iteratively tries to subdivide the parent half-basins into smaller half-basin children and accepts or refuses the new partition under a rule-based criterion. This criterion comprises two distinct aspects. First, the homogeneity of terrain aspect within a SU is controlled through a parameter defined as *circular variance*. Second, the acceptable size of the SU is determined as a function of *flow accumulation* (FA, Jenson and Domingue, 1988) *threshold* and *minimum SU area*. Finally the units below a given extent (or *cleansize*) are dissolved by automatically merging them to the adjacent units.

The final parameterization of *r.slopeunits* has been chosen among 9 possible combinations of circular variance (0.2, 0.4, 0.6), flow accumulation threshold in m^2 ($100 \cdot 10^3$, $250 \cdot 10^3$, $500 \cdot 10^3$) keeping the minimum area and the *cleansize* parameters as

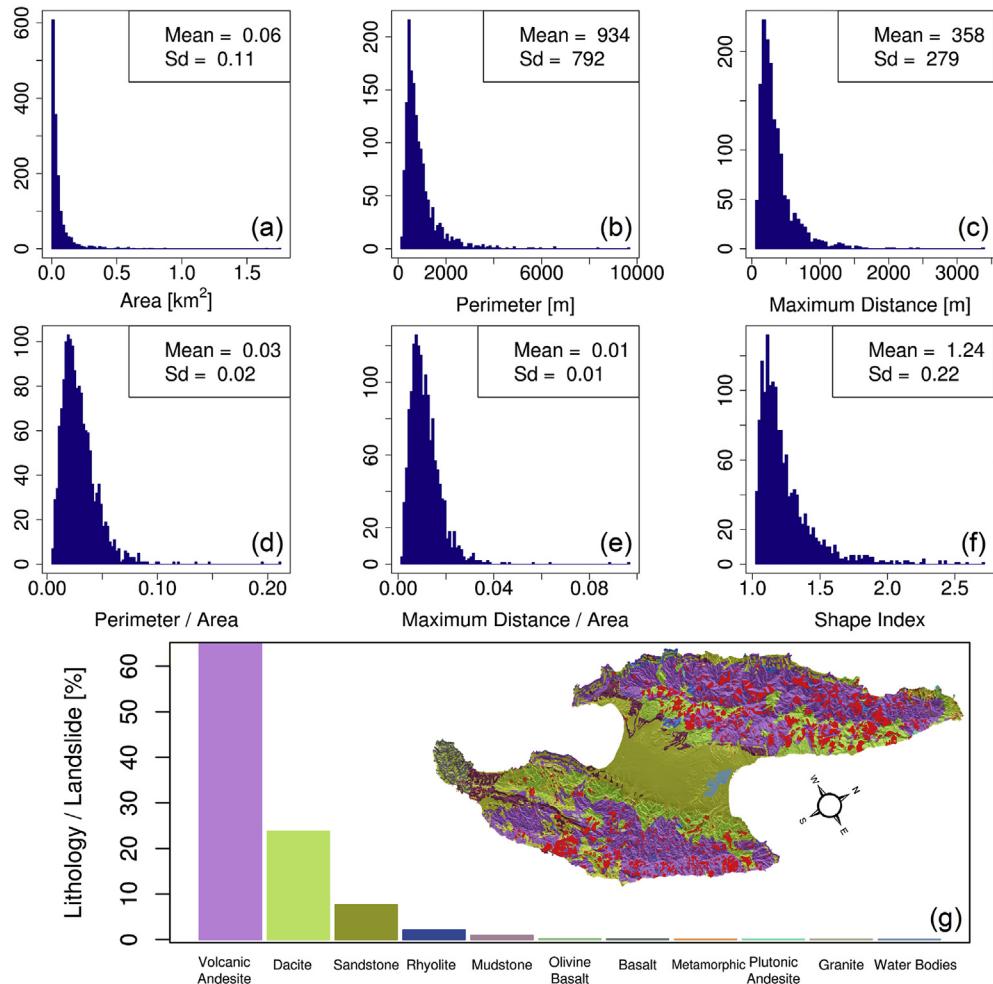


Fig. 3. Landslide characteristics: area (a), perimeter (b), maximum distance within a polygon (c), perimeter to area ratio (d), maximum distance to area ratio (e) shape Index (f) computed following Forman and Gordon (1986) and landslide to lithotype surfaces ratio (g) whose extent is shown in the small map and over imposed with red landslides. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

constant ($50 \cdot 10^3$ and $25 \cdot 10^3$ m², respectively). Fig. SM1 shows the aforementioned combinations, including the selected one which is computed with a FA threshold of $100 \cdot 10^3$ m² and a circular variance of 0.2. The final parameterization has been decided on a qualitative level by comparing the partition to the underlying slope aspect which resulted into 8239 SUs. One of the advantages of adopting slope units is that the high dimensionality of grid-based prediction is summarized from million pixels to thousands SUs thus reducing the associated computational burden.

We extract the Landslide Identification Points (LIPs, hereafter) by calculating the centroid of each of the 1538 mass movements in the Island. The LIP calculation must follow a geomorphological criterion which for the case of deep-seated landslides is commonly adopted to be the centroid (Davis and Sims, 2013; Oliveira et al., 2015; Xu et al., 2012). However, other landslide classes require different representation approaches (Hussin et al., 2016; Lombardo et al., 2014; Süzen and Doyuran, 2004). We then intersect the LIPs with the SU and assign them with a positive status. Therefore, each landslide belongs to one and only one SU. Conversely, every SU that has not crossed a LIP within its boundary has been assigned with a negative status. This operation ensures the creation of a SU-based dichotomous dependent variable expressing landslide presences/absences across the island.

2.3.2. Independent variables

We represent the predictor distribution within a given SU through quantiles and use them as predictors. In particular, we compute for each SU the p -quantiles with probabilities $p = 0.05, 0.10, \dots, 0.95$. This should provide a near-complete description of the distribution of each predictor within SUs, but induces high-multicollinearity which we deal with using a LASSO penalty; see Section 2.4 and Supplementary material for details. In terms of predictor choice, we adopt part of the covariates which were used by Dou et al. (2015a) to explain the landslide occurrences in the same area. This is done to enable comparisons across different studies as the landslide data is the same. In particular, we adopt the following covariates: i) Elevation (10 m DEM), ii and iii) Eastness and Northness (Leempoel et al., 2015), iv and v) Plan and Profile Curvature (Heerdegen and Beran, 1982), vi) Distance to Faults (Pourghasemi et al., 2012), vii) Distance to Geological Boundaries (Dou et al., 2015a), viii) Relative Slope Position (Böhner and Selige, 2006), xi) Slope (Zevenbergen and Thorne, 1987), x) Topographic Wetness Index (Beven and Kirkby (1979), TWI, hereafter). Most of the covariates are derived from the DEM obtained from the Geospatial Information Authority of Japan, GSI (<https://www.hcc.co.jp/work/gismap/youtobetsu/terrain.html>). The actual data provider is Hokkaido-chizu Co. Ltd., a private company that

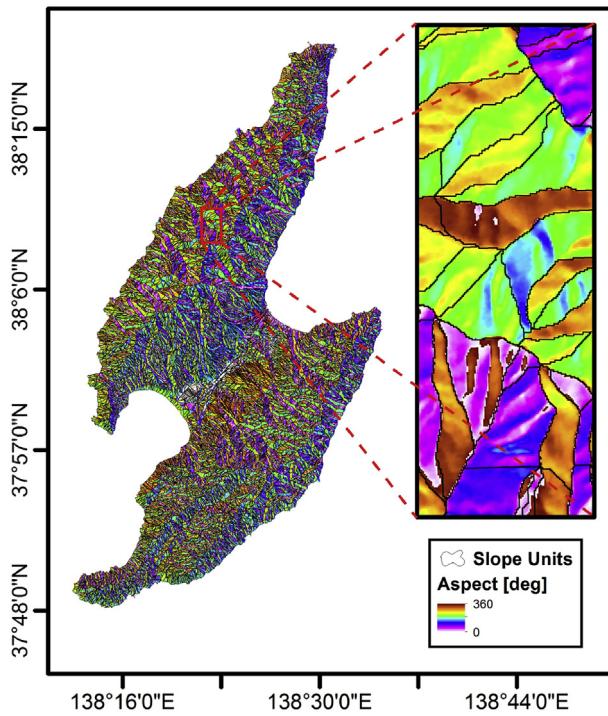


Fig. 4. Slope-unit subdivision of Sado Island is underlain by the slope aspect. The right sub-panel shows the detail for a small region. The legend is valid for the whole island and the zoomed area.

reports a 0.1 m vertical resolution for the whole Japan. Furthermore, we have included the outcropping lithology (scale 1:200,000) produced by the GeoNavi of the Geological Survey of Japan (GSJ). For the latter, we compute the proportion between each lithotype outcropping extent within a given SU and the full extent of the SU itself. This procedure differs from the common practice of getting the predominant lithotype only within a SU (e.g. Tian et al., 2010) and provides a transformation from the original categorical variable to its continuous counterpart expressed through the lithotype to SU surface-ratio. Finally, we have also computed two morphometric indices: the maximum possible distance in a given SU divided by the square root of the area within a given SU ($\max D/\sqrt{A}$, hereafter) and the perimeter divided by the square root of the area within a given SU (P/\sqrt{A} , hereafter), which are measures of roundness of each SU. These indices are computed using the Polygon Shape Indices Module available in SAGA GIS (Conrad et al., 2015). Overall, from an initial set of 13 predictors we generated 273 of them. This number is equal to 13 covariates times 20 (19 quantiles + standard deviation) plus 2 morphometric indices and 11 out of 12 lithology to SU ratios. We excluded the ratio of the most representative lithotype in the island, as it can be expressed as a linear combination of the other lithotype ratios, and this avoids producing a rank-deficient design matrix. In the next section, we discuss how to assess the importance of each predictor and how to perform variable selection, in order to get back to a reasonable and interpretable number of significant and relevant predictors. The physical meaning and interpretation of these covariates may be summarised as follows: i) slope orientation and subsequent potential evapotranspiration effects through Eastness and Northness. These are the continuous counterpart of the Aspect which, being a cyclic number has historically been used as a categorical covariate. Recent contributions (e.g., Steger et al., 2016) are using those by respectively calculating $\sin(\text{Aspect})$ and $\cos(\text{Aspect})$ to produce continuous covariates whose domain are bounded between -1 and

1. For Eastness values equal to 1 indicate exposition facing East whereas -1 faces West. In analogy, for Northness values equal to 1 are facing North and -1 are exposed to the South. The elevation is typically used as ii) a proxy for rainfall intensities while Plan and Profile curvature are interpretable for iii) their influence in directing overland water flows (Ohlmacher, 2007). Distance to Fault indicates iv) pre-existing weakness planes whereas Distance to Geological Boundaries may complement this structural information by indicating v) contacts between lithotypes with different geotechnical properties. Relative Slope Position represent the continuous counterpart to the more general Landform Classification, by expressing vi) pixel positions along the landscape with values equal to 0 for lowland areas and values equal to 1 for ridges. Finally, Topographic Wetness Index expresses vii) the tendency of a given pixel to retain water and subsequently increase pore water pressures.

2.4. LASSO-penalized Generalized Linear Model

In terms of modelling, our goal is to link the dichotomous dependent variable expressing landslide presences/absences with the set of predictors (or independent variables) described in Section 2.3.2. A natural framework for this is provided by logistic regression models, which are constructed using a function that links the probabilities of the binary responses (i.e., the probabilities of landslide presence or absence) with a linear function of the predictors. These models are framed in terms of the generalized linear model (GLM) family, and are able to assess the strength of association between a particular predictor and a landslide occurrence, whilst adjusting for all the other predictors. Details regarding model formulation and estimation can be found in the [Supplementary Material](#).

To make sense of the results obtained from the fitted landslide susceptibility model and to avoid statistical complexities, we need to find an interpretable subset of predictors that are associated with a landslide event. In other words, we need a method to reduce the large number of highly correlated predictors without losing parameter interpretability. Although stepwise variable selection has been a very popular technique to face this issue (see Cama et al. (2017)), several authors (see for instance Copas and Long (1991); Derksen and Keselman (1992); Harrell (2015)) have reported important statistical deficiencies related to the correct distribution of the test statistics, size of the p -values, bias of estimates and standard errors, among other problems. To avoid these issues, we choose to work with the LASSO (Least Shrinkage and Selection Operator) proposed by Tibshirani (1996), designed to deal with multicollinearity in the estimation of large generalized linear models. The LASSO improves prediction error by appropriately shrinking (or penalizing) large regression coefficients towards zero in order to reduce overfitting, while performing simultaneous variable selection and parameter estimation with any generalized linear regression model. Details regarding the LASSO penalization applied to logistic regression can be found in the [Supplementary Material](#).

Regarding the practical implementation of these methods, we have structured a code (called LUDARA) to fit and validate our model parallelizing over a certain number of replicates N . Specifically, our code (which can be found in the [Supplementary Material](#)) proceeds as follows: i) it randomly extracts 75% of landslide presences in the whole region, as well as the same number of landslide absences according to a balanced sampling strategy. The logistic regression model is then fitted using a LASSO penalty optimally chosen by cross-validation. ii) the code then stores the predictor IDs and corresponding coefficients. iii) the complementary 25% of presences is then used for validation together with an

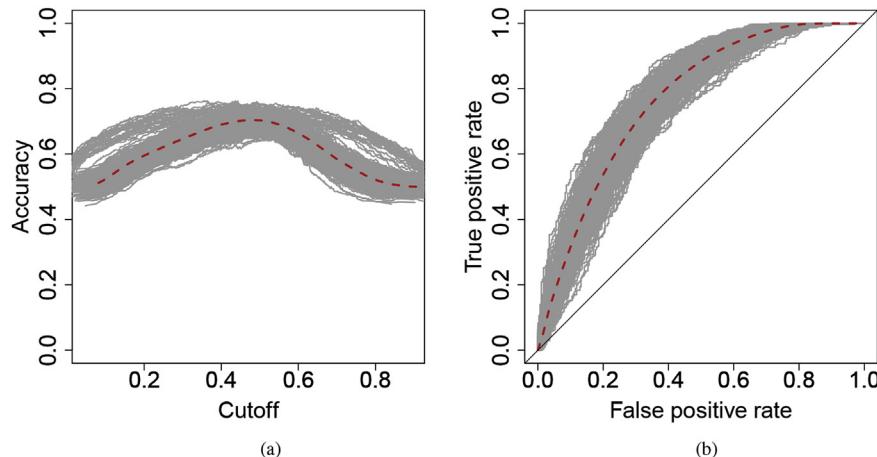


Fig. 5. Accuracy test (a) and Receiver Operating Characteristic curves (b) for each of the 500 replicates. Red lines represent the corresponding means. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

equal number of (randomly chosen) absences; using the validation subset, iv) the Accuracy or (*True Positives + True Negatives*)/Total Population is computed varying the probability cutoff value that separates landslide from no-landslide cases; v) the Receiver Operating Characteristic (ROC) curve and the AUC (area under the ROC curve) are computed; vi) the whole procedure is repeated $N = 500$ times and therefore produces N different models with their corresponding parameter estimates, ROC curves and AUCs. The predictor importance may be assessed by counting the actual selections of the 273 predictors across the N models. Mean fitted probabilities, mean ROC curve and mean AUC are obtained by averaging over all the N replicates; vii) the accuracy is plotted against the varying cutoffs and the maximum accuracy value of the average resulting curves is then adopted to separate stable from unstable landslide conditions. Step vii) is implemented to quantitatively determine the best probability cutoff as several authors (Felcísmo et al., 2013; Frattini et al., 2010; Heckmann et al., 2014) have pointed out that the cutoff choice is not unique and may change with the sampling scheme, method or dataset. Details regarding the model performances are given in Sections 3 and 4.

3. Results

Fig. 5(a) shows the results of the accuracy test across 500 replicates where the average maximum accuracy of 0.71 has been reached at approximately the standard cutoff of 0.5; this value was fixed for the rest of the analyses for confusion matrices (e.g. Lombardo et al., 2016b). Fig. 5(b) shows the validation ROC curves across 500 replicates. The corresponding average AUC of 0.78 is associated to a standard deviation of 0.03. This attests for an almost excellent prediction skill (Hosmer and Lemeshow, 2000) together with a strong stability of the model throughout the cross-validation procedure. Once the quality of the models is confirmed in terms of overall prediction skill and associated variance, the predictor behavior has been investigated to infer the geomorphological reasonability of the results. In particular, due to the combined random data set extraction, high dimensionality of the predictor hyperspace and the LASSO-based variable selection, the list of significant covariates changes from each replicate. In order to get a unique interpretable set of predictors, we have selected the covariates that are present within each of the 500 replicates more than 80% of the time. Fig. 6 plots the percentage of selection for each predictor.

The eight covariates above the red line are sorted in order of

importance and are listed as follows: i) maxD/sqrtA, ii) Sandstone/SU, iii) Elevation's 95th percentile, iv) TWI's 95th percentile, v) SD of Elevation, vi) Plan Curvature's 5th percentile, vii) Relative Slope Position's 5th percentile, viii) Eastness' 95th percentile. The effects of these covariates on the final susceptibility are assessed analysing the regression β coefficient values (Cama et al., 2015, 2016) together with response plots (Pourghasemi and Rossi, 2016; Schillaci et al., 2017a,b; Tziritis and Lombardo, 2016). This is shown in Fig. 7 where the distribution of the 500 coefficients is summarized. The mean landslide susceptibility (computed as the average susceptibility from the 500 replicates) is plotted against each covariate in Fig. 8. This approach complements the analyses of the coefficients providing a more dynamic perspective of variable-interaction effects on the susceptibility. Moreover, the mean and standard deviation of the estimated susceptibility for each SU are mapped in Fig. 9(a) and Fig. 9(b). These two elements allow for visualizing the prediction and for completing the evaluation on the model performances. This evaluation consists of computing the model error (Lombardo et al., 2016a; Rossi et al., 2010) where the mean susceptibility value is plotted against twice the standard deviation. Fig. 10 presents the model error in a density plot with an upper bound of the 95% confidence interval coinciding with 0.2.

4. Discussion

We tested our code, LUDARA, within the Sado Island where a well-tested inventory of deep-seated landslides was made

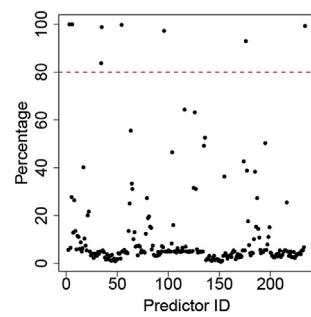


Fig. 6. Variable importance measured by percentage of selection over the 500 replicates. The red horizontal dashed line represents the 80% variable-inclusion threshold. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

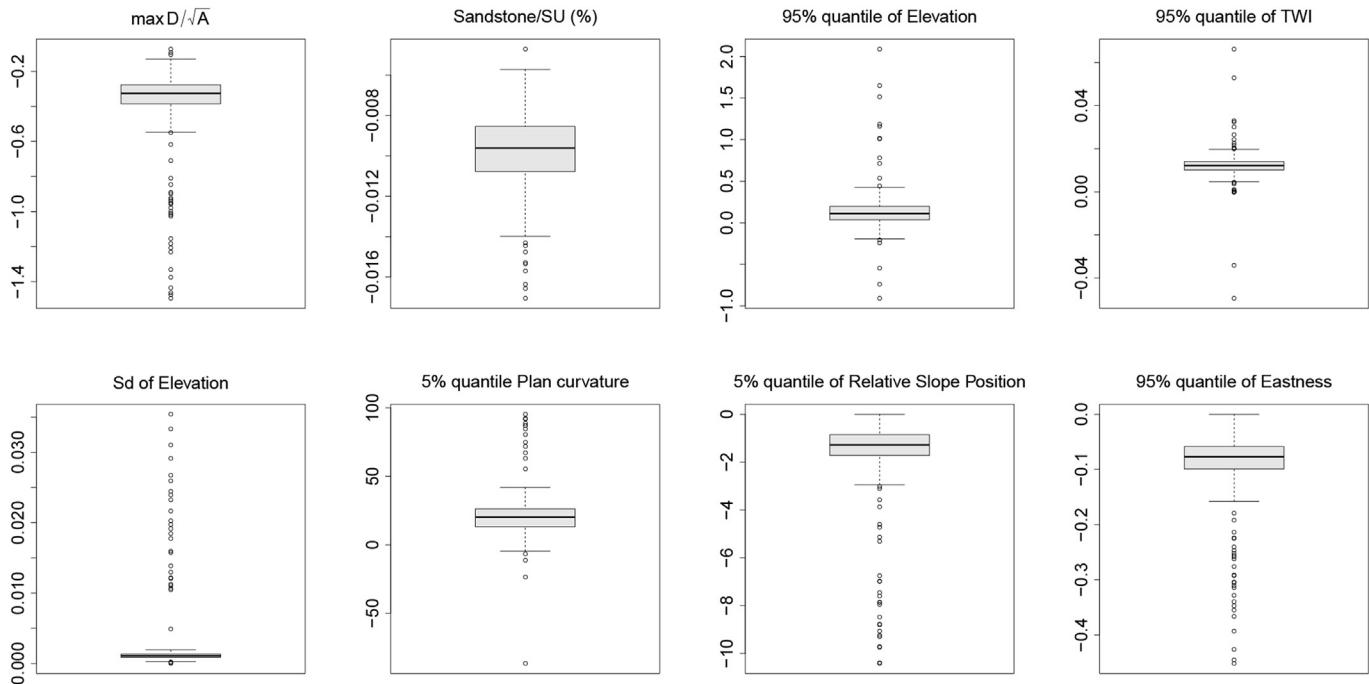


Fig. 7. Coefficient boxplots of the eight most selected covariates, based on 500 replicates.

available. The model performances were evaluated at different stages. The first stage implemented an accuracy test to select the best probability cutoff for separating stable from unstable conditions in the probability spectrum. This was done by maximizing the accuracy as a function of a moving cutoff window. The average maximum accuracy across replicates was 0.71 at a probability cutoff value of 0.5 which is also the commonly adopted threshold. The same threshold was subsequently applied to the rest of the metric

calculations (TP, FP, TN, FN). In addition, the ROC curves and corresponding average AUC value of 0.78 confirmed an almost excellent prediction together with low associated variance.

The geomorphological reasonability of the model was inferred on those covariates which were selected more than 400 times out of the total 500 replicates. Fig. 8 summarizes their coefficients during the modeling phase while Fig. 7 shows the predictor response over the fitted probabilities for each of the SU in the

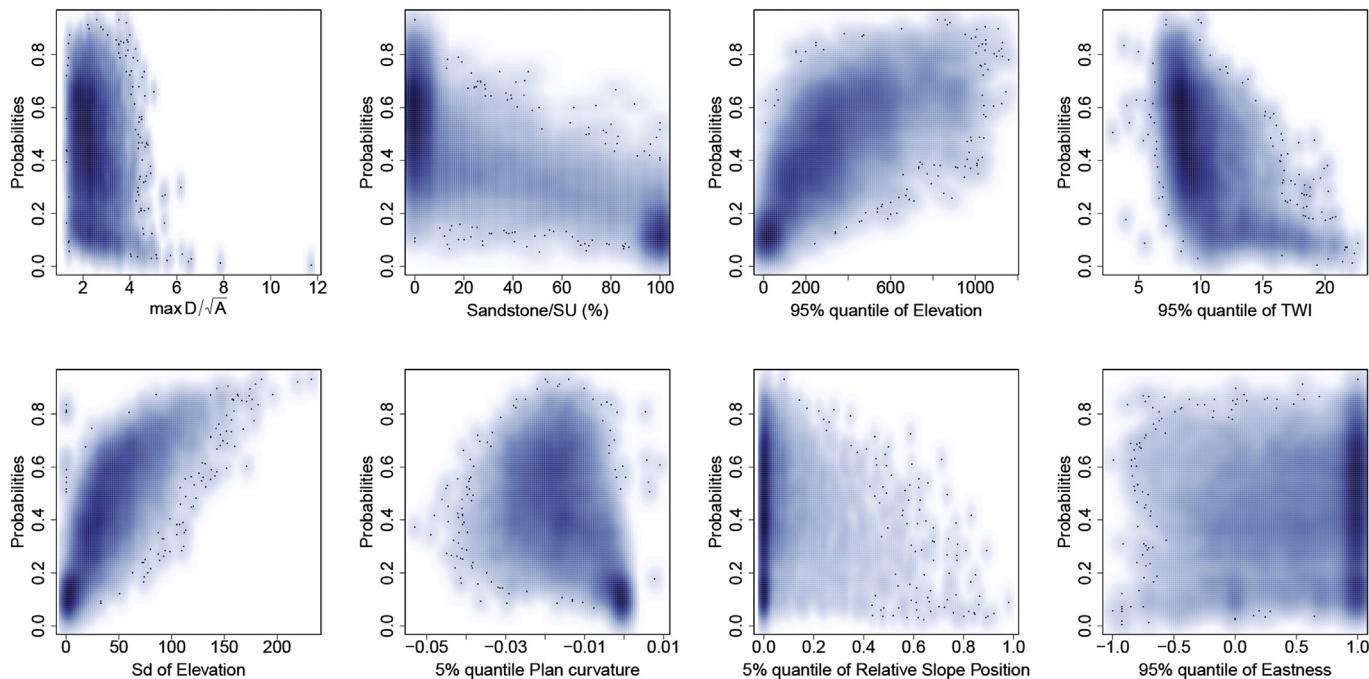


Fig. 8. Response density plot of the eight most selected covariates based on 500 replicates. The figure was generated by applying a two-dimensional (kernel density) smoothing, over imposing the outliers as black dots.

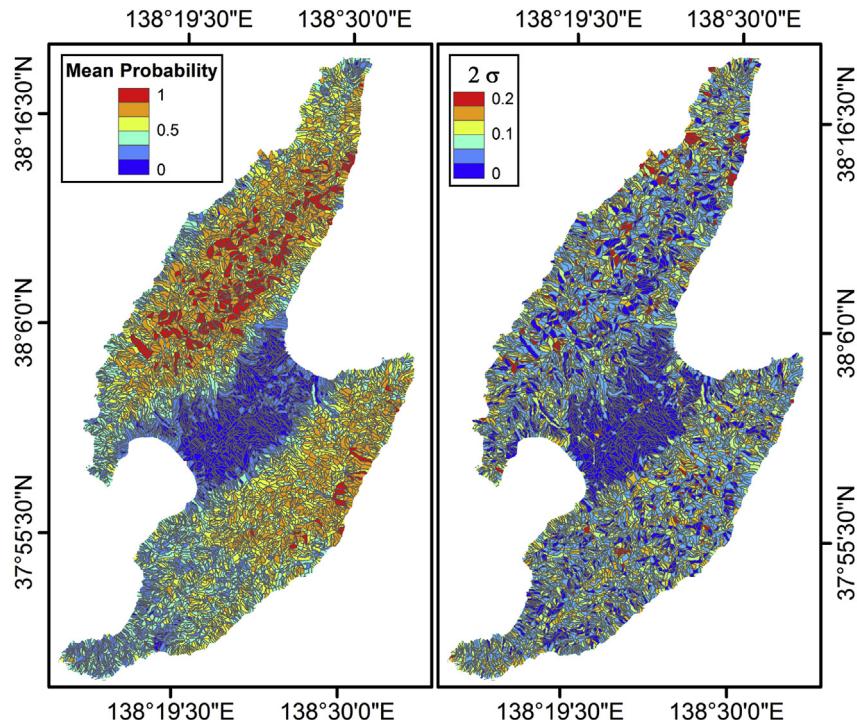


Fig. 9. Mean (left panel) and twice standard deviation (right panel) susceptibility maps. Palettes are divided in equal intervals between the respective minima and maxima.

island. MaxD/\sqrt{A} appeared to be the most significant covariate. This shape index carries the information on the elongation of a given SU which decreases as the ratio increases. The sign of the corresponding β coefficient is negatively associated with landslide presences (median $\beta = -0.33$) indicating that the less elongated the SU the more unstable it would be. This can be interpreted both as a classification effect, as most of the landslides in the area are large, or as a lower shear resistance along the sliding surface against greater unstable volumes. The same relation can be seen on the response plots where the probability rapidly decays as the maxD/\sqrt{A} increases.

The Sandstone to SU areal ratio similarly shows a negative effect on landslide occurrences (median $\beta = -0.01$) which is confirmed over the response with two evident clusters, the main one being strictly unstable where Sandstone/SU is absent and a second cluster with very low probabilities at high Sandstone proportions. It is worth noting that the Sandstone was shown in Fig. 3(g) to be the

third most representative outcropping lithotype for the landslides. Despite this, the regression β coefficient was estimated to be negative. This can depend on the limited thickness of these sandstone deposits when compared to the sliding surface of a deep-seated landslide together with the prevalent presence of this covariate along the central graben where the topography is flat. The Elevation's 95th percentile has a positive effect on instabilities (median $\beta = 0.11$) which results in a rapid increase in the SU susceptibility. This can be explained by the fact that this covariate represents the highest elevation values within a SU which can also be translated into a greater exposition to rainfall discharges, which in turn destabilize the slope unit.

The TWI's 95th percentile contributes to SU instability (median $\beta = 0.01$). The corresponding response plot distribution shows a peak at 8 where the maximum instability is reached. This distribution can be interpreted as the destabilizing effect of water infiltration where the landslide susceptibility is high while the decay to stable conditions can be due to such elevated TWI values that would only correspond to the flat regions along the central plain.

The standard deviation of the Elevation was also included, being characterized by a positive β coefficient of 0.001 and a clear trend in the response plot. This variable can be seen as a proxy for topographic roughness which is known to affect the stability of a slope (e.g. Mahalingam et al., 2016).

The Plan Curvature's 5th percentile is also positively linked to landslide presences (median $\beta = 20.10$). However, as the predictor itself is negative at low quantiles, the response plot shows a probability decrease down to a stable cluster at around 0. This can be interpreted as the concave focusing effect of topography on overland water flows (Ohlmacher, 2007).

The Relative Slope Position's 5th percentile is negative (median $\beta = -1.28$) and clearly reflects the stable conditions of lower portions of the landscape. However, the signal in the response plot is noisy and difficult to be distinguished probably due to interaction effects with other variables.

Similarly, the Eastness' 95th percentile is easily interpretable on

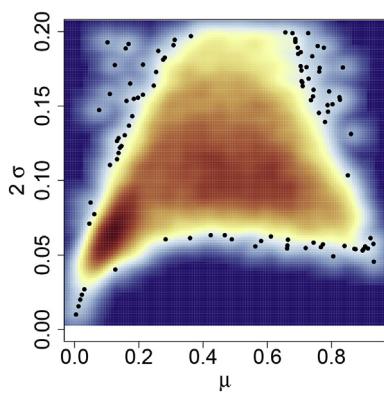


Fig. 10. Model error plot, mean susceptibility (μ) versus 95% confidence prediction interval (2σ).

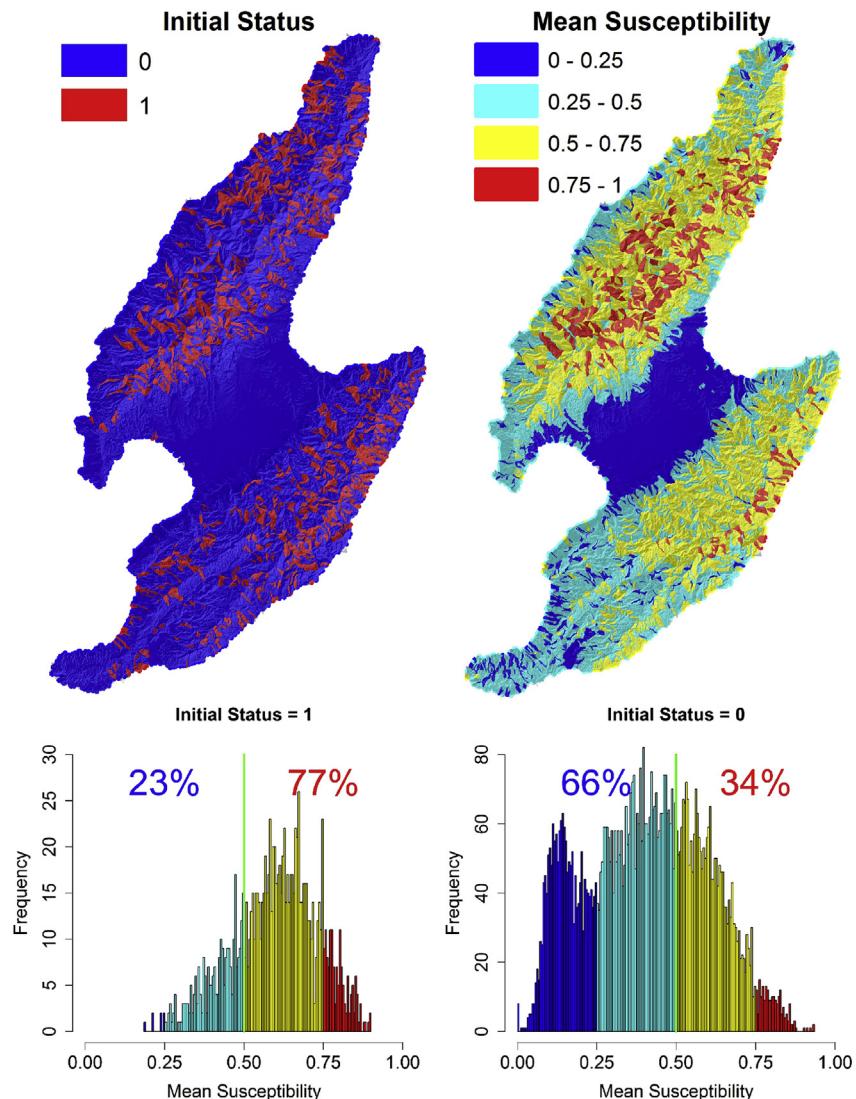


Fig. 11. Original data (top left panel) and mean susceptibility map (top right panel). This information is combined in the lower panels where the assigned probabilities are shown for unstable and stable slope units. The green line represents the probability cutoff at 0.5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the regression β coefficient boxplot (median $\beta = -0.08$) but masked within the response plot where a unit increase of the Eastness does not reflect a decrease in landslide susceptibility.

These results agree with those described by Dou et al. (2015a,b) where the Slope aspect was strongly correlated to landslide presences in the East and South-East direction. Another agreement was also found with the Plan Curvature and Elevation, these being identified as most relevant contributors in the two aforementioned researches. Furthermore, the analyses were completed by regionalizing the prediction over the Sado Island. Fig. 9 shows the mean and twice the standard deviation of the susceptibility estimates. The former clearly recognizes a greater susceptibility to the north ridge which was also reflected on the landslide inventory. This is an interesting observation as the susceptibility map was built to represent a global model for the island without carrying the specific spatial locations of the landslides. Despite this, the model was able to capture the spatial pattern from the covariates. In addition, the Northern region was depicted with a similar susceptibility distribution with respect to the one shown in Dou et al. (2015a,b) and being constructed upon a grid-based mapping unit.

With respect to the standard deviation map, the low variability of the estimates confirm the stability and thus the reliability across replicates. No evident patterns arose from this map with high standard deviation values scattered across the island likely due to very local features. From a modeling perspective, this random distribution of the standard deviation is a good indication as high values strongly clustered in space would have suggested a limited capacity of the global susceptibility model to fit specific portions of the island. Level 4 validation procedures (Guzzetti et al., 2006b) were completed by crossing the two aforementioned maps to generate the error plot for Sado Island. It is worth mentioning that this plot should ideally produce a bell shape which can be seen in Fig. 10 with the exception of a cluster in the left tail due to a higher density of stable SUs in the center of the island.

Ultimately, one of the issues in landslide prediction is that susceptibility maps sometimes only mark high probabilities where landslides are already present but fail to highlight locations with similar instabilities. For such cases, overfitting is generally the cause (Braun et al., 2015) and landslide susceptibility hardly provides new information for the given study area. We here check for

potential overfitting effects by assessing the predicted probability assigned at SUs and comparing it with respect to the original status (Fig. 11). The test highlights that 77% of the original landslide cases are correctly predicted. However, 34% of the originally stable SUs are assigned with high landslide susceptibilities, indicating potential instability issues and providing useful information for risk prevention.

5. Conclusions

The assumption behind this contribution was due to a literature review on Slope Unit-based landslide susceptibility where no papers had considered the overall predictor distribution within a given SU. We attempted to sample the distributions in terms of quantiles and implemented a LASSO penalization together with logistic regression to infer the most significant covariates keeping the models as skillful as possible. The SUs were calculated using the most recent development in the specific literature thanks to r.slopeunits. This code is freely accessible and provides unambiguous parameters for the SU delineation. Our LUDARA code was implemented to generate 500 replicates which produced excellent performances and low variances. The same performances were evaluated on the basis of common metrics with the difference that, instead of blindly applying a cutoff value between stable and unstable conditions, we objectively determined the cutoff value maximizing the accuracy as a function of a moving threshold. The value actually coincided with the common 0.5 probability cutoff.

The most original findings are to be related to the predictors. No predictor was selected to be the median within the SU, which could simplistically be approximated to the commonly adopted mean. On the contrary, the most relevant covariates either fell on 5th or 95th percentile which supports the initial assumption that using just the mean value may only provide part of the instability information. This is an interesting indication as pixel-based models would overlook these instability values unless the polygon to point conversion would generate a point in a high or low quantile pixel. This may not be the case and it strongly depends on the assumption by each researcher when approximating the landslide scar to a single position. Conversely, the advantage of SU-based model would certainly include high and low quantile pixels modelling instabilities as an areal rather than a point process.

The variable selection was also implemented differently from the general stepwise procedure. Another difference from the available literature consists of the inspection of the covariate importance (percentage of selection), role (static view of the beta coefficients) and behaviour (dynamic perspective of the response plot). These last two tools allowed for recognizing the: i) Elongation, ii) proportion of the Sandstone, iii) extreme values of the Elevation, iv) TWI, v) Eastness, vi) Plan Curvature and vii) Relative Slope Position together with the viii) standard deviation of the elevation as the most relevant contributors to the final susceptibility assessment.

The subsequent susceptibility map produced realistic spatial patterns which replicate the greater proneness to landsliding in the north ridge without having implemented any strict spatial criteria.

Finally, our LUDARA code, was scripted into R to make it freely accessible to the community. LUDARA is parallelized to speed up the computation time even on supercomputers. However, it can easily be run on common multicore machines with lower capacities.

Taking a step back and looking at the overall picture, we suggest the implementation of a similar approach for any terrain-unit based susceptibility study, where any subdivision of the geographic space results in a polygonal segmentation. There is no guarantee that common mean-based models would produce similar spatial

predictive patterns to those obtained by modelling the whole distribution. As a consequence, we recommend the data themselves to guide the actual variable selection process and in turn the optimal predictive map.

This suggestion arises from considering the end users involved with the risk management who base their decision on the outcome of modelling procedures. The investments on slope stabilization are often tailored to the causative factors that play a major role in the susceptibility and a biased judgement due to a partial investigation may affect the success of the stabilization itself. In other words, the a priori use of the mean only may introduce limitations on the susceptibility map whereas a quantitative selection of the best covariate would produce objective susceptibility maps to operate upon for risk assessment.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.envsoft.2017.08.003>.

References

- Abdulwahid, W.M., Pradhan, B., 2016. Landslide vulnerability and risk assessment for multi-hazard scenarios using airborne laser scanning data (LiDAR). *Landslides* 1–20.
- Aleotti, P., Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. *Bull. Eng. Geol. Environ.* 58 (1), 21–44.
- Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., Guzzetti, F., 2016. Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geosci. Model Dev.* 9 (11), 3975–3991.
- An, H., Viet, T.T., Lee, G., Kim, Y., Kim, M., Noh, S., Noh, J., 2016. Development of time-variant landslide-prediction software considering three-dimensional subsurface unsaturated flow. *Environ. Model. Softw.* 85, 172–183.
- Arnone, E., Francipane, A., Scarbaci, A., Puglisi, C., Noto, L., 2016. Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. *Environ. Model. Softw.* 84, 467–481.
- Ayalew, L., Yamagishi, H., Marui, H., Kanno, T., 2005a. Landslides in Sado Island of Japan: Part I. Case studies, monitoring techniques and environmental considerations. *Eng. Geol.* 81 (4), 419–431.
- Ayalew, L., Yamagishi, H., Marui, H., Kanno, T., 2005b. Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Eng. Geol.* 81 (4), 432–445.
- Baeza, C., Corominas, J., 2001. Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. *Earth Surf. Process. Landforms* 26 (12), 1251–1263.
- Beguería, S., Lorente, A., 2007. Landslide Hazard Mapping by Multivariate Statistics: Comparison of Methods and Case Study in the Spanish Pyrenees.
- Beven, K., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sci. J.* 24 (1), 43–69.
- Böhner, J., Selige, T., 2006. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. *Göttinger Geogr. Abh.* 115, 13–28.
- Braun, A., Fernandez-Steger, T., Havenith, H.-B., Torgoev, A., 2015. Landslide Susceptibility Mapping with Data Mining Methods—a Case Study from Maily-Say, Kyrgyzstan. Springer International Publishing, Cham, pp. 995–998.
- Brenning, A., 2005. Spatial prediction models for landslide hazards: review, comparison and evaluation. *Nat. Hazards Earth Syst. Sci.* 5 (6), 853–862.
- Calvello, M., Cascini, L., Mastrianni, S., 2013. Landslide zoning over large areas from a sample inventory by means of scale-dependent terrain units. *Geomorphology* 182, 33–48.
- Cama, M., Lombardo, L., Conoscenti, C., Agnesi, V., Rotigliano, E., 2015. Predicting storm-triggered debris flow events: application to the 2009 Ionian Peloritan disaster (Sicily, Italy). *Nat. Hazards Earth Syst. Sci.* 15 (8), 1785–1806.
- Cama, M., Conoscenti, C., Lombardo, L., Rotigliano, E., 2016. Exploring relationships between grid cell size and accuracy for debris-flow susceptibility models: a test in the Giampilieri catchment (Sicily, Italy). *Environ. Earth Sci.* 75 (3), 1–21.
- Cama, M., Lombardo, L., Conoscenti, C., Rotigliano, E., 2017. Improving transferability strategies for debris flow susceptibility assessment: application to the Saponara and Itala catchments (Messina, Italy). *Geomorphology* 288, 52–65.
- Capitani, M., Ribolini, A., Federici, P., 2013. Influence of deep-seated gravitational slope deformations on landslide distributions: a statistical approach. *Geomorphology* 201, 127–134.
- Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., Reichenbach, P., 1991. Gis techniques and statistical models in evaluating landslide hazard. *Earth Surf. Process. Landforms* 16 (5), 427–445.
- Carrara, A., Cardinali, M., Guzzetti, F., Reichenbach, P., 1995. Gis Technology in Mapping Landslide Hazard. Springer Netherlands, Dordrecht, pp. 135–175.

- Carrara, A., Crosta, G., Frattini, P., 2008. Comparing models of debris-flow susceptibility in the alpine environment. *Geomorphology* 94 (3), 353–378.
- Chung, C.-J.F., Fabbri, A.G., Van Westen, C.J., 1995. Multivariate regression analysis for landslide hazard zonation. In: *Geographical Information Systems in Assessing Natural Hazards*. Springer, pp. 107–133.
- Chung, C.-J.F., Fabbri, A.G., 1999. Probabilistic prediction models for landslide hazard mapping. *Photogrammetric Eng. remote Sens.* 65 (12), 1389–1399.
- Conoscenti, C., Rotigliano, E., Cama, M., Caraballo-Arias, N.A., Lombardo, L., Agnesi, V., 2016. Exploring the effect of absence selection on landslide susceptibility models: a case study in Sicily, Italy. *Geomorphology* 261, 222–235.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (saga) v. 2.1.4. *Geosci. Model Dev.* 8 (7), 1991–2007.
- Copas, J., Long, T., 1991. Estimating the residual variance in orthogonal regression with variable selection. *Statistician* 51–59.
- Felicísimo, Á.M., Cuartero, A., Remondo, J., Quirós, E., 2013. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides* 10 (2), 175–189.
- Davis, J.D., Sims, S.M., 2013. Physical and maximum entropy models applied to inventories of hill slope sediment sources. *J. Soils Sediments* 13 (10), 1784–1801.
- Derkzen, S., Keselman, H., 1992. Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *Br. J. Math. Stat. Psychol.* 45 (2), 265–282.
- Dou, J., Bui, D.T., Yunus, A.P., Jia, K., Song, X., Revhaug, I., Xia, H., Zhu, Z., 2015a. Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. *PloS One* 10 (7), e0133262.
- Dou, J., Yamagishi, H., Pourghasemi, H.R., Yunus, A.P., Song, X., Xu, Y., Zhu, Z., 2015b. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. *Nat. Hazards* 78 (3), 1749–1776.
- Erenler, A., Düzgün, H., 2012. Landslide susceptibility assessment: what are the effects of mapping unit and mapping method? *Environ. Earth Sci.* 66 (3), 859–877.
- Forman, R., Gordon, M., 1986. *Landscape Ecology*, John.
- Frattini, P., Crosta, G., Carrara, A., 2010. Techniques for evaluating the performance of landslide susceptibility models. *Eng. Geol.* 111 (1), 62–72.
- Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., Reichenbach, P., 2008. Comparing landslide inventory maps. *Geomorphology* 94 (3), 268–289.
- Ganzawa, Y., 1983. Green tuff movement defined by fission track ages of igneous rocks.(2). *Chishitsugaku Zasshi* 89 (5), 271–286.
- Goetz, J., Brenning, A., Petschko, H., Leopold, P., 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosciences* 81, 1–11.
- Guzzetti, F., Reichenbach, P., 1994. Towards a definition of topographic divisions for Italy. *Geomorphology* 11 (1), 57–74.
- Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P., 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31 (1), 181–216.
- Guzzetti, F., Cardinali, M., Reichenbach, P., Carrara, A., 2000. Comparing landslide maps: a case study in the upper Tiber River Basin, central Italy. *Environ. Manag.* 25 (3), 247–263.
- Guzzetti, F., Galli, M., Reichenbach, P., Ardizzone, F., Cardinali, M., 2006a. Landslide hazard assessment in the Collazzone area, Umbria, Central Italy. *Nat. Hazards Earth Syst. Sci.* 6 (1), 115–131.
- Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Galli, M., 2006b. Estimating the quality of landslide susceptibility models. *Geomorphology* 81 (1), 166–184.
- Hansen, A., 1984. *Landslide Hazard Analysis. Slope Instability*. Wiley, New York, pp. 523–602.
- Harrell, F., 2015. *Regression Modeling Strategies: with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer, New York.
- Heckmann, T., Gegg, K., Gegg, A., Becht, M., 2014. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. *Nat. Hazards Earth Syst. Sci.* 14 (2), 259–278.
- Heerdegen, R.G., Beran, M.A., 1982. Quantifying source areas through land surface curvature and shape. *J. Hydrology* 57 (3–4), 359–373.
- Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression*, second ed.
- Huabin, W., Gangjun, L., Weiye, X., Gonghui, W., 2005. Gis-based landslide hazard assessment: an overview. *Prog. Phys. Geogr.* 29 (4), 548–567.
- Hussin, H.Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., Bălteanu, D., 2016. Different landslide sampling strategies in a grid-based bivariate statistical susceptibility model. *Geomorphology* 253, 508–523.
- Iovine, G., Di Gregorio, S., Sheridan, M., Miyamoto, H., 2007. Modelling, computer-assisted simulations, and mapping of dangerous phenomena for hazard assessment. *Environ. Model. Softw.* 22 (10), 1389–1391.
- Jenson, S.K., Domingue, J.O., 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Eng. remote Sens.* 54 (11), 1593–1600.
- Jolivet, L., Huchon, P., Brun, J.P., Le Pichon, X., Chamot-Rooke, N., Thomas, J.C., 1991. Arc deformation and marginal basin opening: Japan Sea as a case study. *J. Geophys. Res. Solid Earth* 96 (B3), 4367–4384.
- Jolivet, L., Fournier, M., Huchon, P., Rozhdestvenskiy, V.S., Sergeyev, K.F., Orcobin, L.S., 1992. Cenozoic intracontinental dextral motion in the Okhotsk-Japan sea region. *Tectonics* 11 (5), 968–977.
- Leempoel, K., Parisod, C., Geiser, C., Daprà, L., Vittoz, P., Joost, S., 2015. Very high-resolution digital elevation models: are multi-scale derived variables ecologically relevant? *Methods Ecol. Evol.* 6 (12), 1373–1383.
- Lombardo, L., Cama, M., Maerker, M., Rotigliano, E., 2014. A test of transferability for landslides susceptibility models under extreme climatic events: application to the Messina 2009 disaster. *Nat. Hazards* 74 (3), 1951–1989.
- Lombardo, L., Cama, M., Conoscenti, C., Märker, M., Rotigliano, E., 2015. Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: application to the 2009 storm event in Messina (Sicily, southern Italy). *Nat. Hazards* 79 (3), 1621–1648.
- Lombardo, L., Bachofer, F., Cama, M., Märker, M., Rotigliano, E., 2016a. Exploiting Maximum Entropy method and ASTER data for assessing debris flow and debris slide susceptibility for the giampilieri catchment (north-eastern Sicily, Italy). *Earth Surf. Process. Landforms* 41 (12), 1776–1789.
- Lombardo, L., Fibelli, G., Amato, G., Bonasera, M., 2016b. Presence-only approach to assess landslide triggering-thickness susceptibility: a test for the Mili catchment (north-eastern Sicily, Italy). *Nat. Hazards* 84 (1), 565–588.
- Mahalingam, R., Olsen, M.J., O'Banion, M.S., 2016. Evaluation of landslide susceptibility mapping techniques using lidar-derived conditioning factors (Oregon case study). *Geomatics, Nat. Hazards Risk* 1–24.
- McKean, J., Roering, J., 2004. Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. *Geomorphology* 57 (3), 331–351.
- Metz, M., Mitasova, H., Harmon, R., 2011. Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search. *Hydrology Earth Syst. Sci.* 15 (2), 667–678.
- Neteler, M., Mitasova, H., 2013. *Open Source GIS: a GRASS GIS Approach*, vol. 689. Springer Science & Business Media.
- Nichol, J., Wong, M., 2005. Detection and interpretation of landslides using satellite images. *Land Degrad. Dev.* 16 (3), 243–255.
- Ohlmacher, G.C., 2007. Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Eng. Geol.* 91 (2), 117–134.
- Oliveira, S.C., Zézere, J.L., Garcia, R.A., 2015. Structure and characteristics of landslide input data and consequences on landslide susceptibility assessment and prediction capability. In: *Engineering Geology for Society and Territory*, vol. 2. Springer, pp. 189–192.
- Pourghasemi, H.R., Rossi, M., 2016. Landslide susceptibility modeling in a landslide prone area in Mazandaran Province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP methods. *Theor. Appl. Climatol.* 1–25.
- Pourghasemi, H., Pradhan, B., Gokeoglu, C., Moezzi, K.D., 2012. Landslide susceptibility mapping using a spatial multi criteria evaluation model at Haraz Watershed, Iran. In: *Terrigenous Mass Movements*. Springer, pp. 23–49.
- Reichenbach, P., Mondini, A., Rossi, M., et al., 2014. The influence of land use change on landslide susceptibility zonation: the briga catchment test site (Messina, Italy). *Environ. Manag.* 54 (6), 1372–1384.
- Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A.C., Peruccacci, S., 2010. Optimal landslide susceptibility zonation based on multiple forecasts. *Geomorphology* 114 (3), 129–142.
- Rotigliano, E., Cappadonia, C., Conoscenti, C., Costanzo, D., Agnesi, V., 2012. Slope units-based flow susceptibility model: using validation tests to select controlling factors. *Nat. hazards* 61 (1), 143–153.
- Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., Saia, S., 2017a. Spatio-temporal topsoil organic carbon mapping of a semi-arid mediterranean region: the role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. *Sci. Total Environ.* 601, 821–832.
- Schillaci, C., Lombardo, L., Saia, S., Fantappiè, M., Märker, M., Acutis, M., 2017b. Modelling the topsoil carbon stock of agricultural lands with the Stochastic Gradient Treeboost in a semi-arid Mediterranean region. *Geoderma* 286, 35–45.
- Steger, S., Brenning, A., Bell, R.S., Glade, T., 2016. The propagation of inventory-based positional errors into statistical landslide susceptibility models. *Nat. Hazards Earth Syst. Sci.* 16 (12), 2729–2745.
- Süzen, M.L., Doyuran, V., 2004. A comparison of the gis based landslide susceptibility assessment methods: multivariate versus bivariate. *Environ. Geol.* 45 (5), 665–679.
- Süzen, M.L., Kaya, B.S., 2012. Evaluation of environmental parameters in logistic regression models for landslide susceptibility mapping. *Int. J. Digital Earth* 5 (4), 338–355.
- Tarolli, P., Sofia, G., Dalla Fontana, G., 2012. Geomorphic features extraction from high-resolution topography: landslide crowns and bank erosion. *Nat. Hazards* 61 (1), 65–83.
- Tian, Y., Xiao, C., Wu, L., 2010. Slope unit-based landslide susceptibility zonation. In: *2010 18th International Conference on Geoinformatics*. IEEE, pp. 1–5.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* 267–288.
- Tziritis, E., Lombardo, L., 2016. Estimation of intrinsic aquifer vulnerability with index-overlay and statistical methods: the case of eastern Kopaïda, central Greece. *Appl. Water Sci.* 1–15.
- Van Den Eeckhout, M., Reichenbach, P., Guzzetti, F., Rossi, M., Poessens, J., 2009. Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium. *Nat. Hazards Earth Syst. Sci.* 9 (2), 507–521.
- Van Westen, C.J., 2000. The modelling of landslide hazards using GIS. *Surv. Geophys.* 21 (2–3), 241–255.

- Van Westen, C., Van Asch, T.W., Soeters, R., 2006. Landslide hazard and risk zonationwhy is it still so difficult? *Bull. Eng. Geol. Environ.* 65 (2), 167–184.
- Xu, C., Dai, F., Xu, X., Lee, Y.H., 2012. Gis-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* 145, 70–80.
- Yamagishi, H., 2008. GIS mapping of landscape and disasters of Sado Island, Japan. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 37, 1429–1432.
- Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography. *Earth Surf. Process. Landforms* 12 (1), 47–56.
- Zhou, S., Fang, L., Liu, B., 2015. Slope unit-based distribution analysis of landslides triggered by the April 20, 2013, Ms 7.0 Lushan earthquake. *Arabian J. Geosciences* 8 (10), 7855–7868.