Geomorphology 261 (2016) 222–235

Contents lists available at ScienceDirect

Geomorphology

journal homepage: www.elsevier.com/locate/geomorph

Exploring the effect of absence selection on landslide susceptibility models: A case study in Sicily, Italy

Christian Conoscenti [a,*], Edoardo Rotigliano [a], Mariaelena Cama [a], Nathalie Almaru Caraballo-Arias [b], Luigi Lombardo [a], Valerio Agnesi [a]

[a] Department of Earth and Marine Sciences (DISTEM), University of Palermo, Via Archirafi 22, 90123 Palermo, Italy
[b] Department of Agricultural and Forestry Sciences, University of Palermo, Viale delle Scienze, 90128 Palermo, Italy

ARTICLE INFO

Article history:
Received 2 August 2015
Received in revised form 24 November 2015
Accepted 10 March 2016
Available online 11 March 2016

Keywords:
Landslide susceptibility
Google Earth
Geographic Information Systems (GIS)
Multivariate Adaptive Regression Splines (MARS)

ABSTRACT

A statistical approach was employed to model the spatial distribution of rainfall-triggered landslides in two areas in Sicily (Italy) that occurred during the winter of 2004–2005. The investigated areas are located within the Belice River basin and extend for 38.5 and 10.3 km^2, respectively. A landslide inventory was established for both areas using two Google Earth images taken on October 25th 2004 and on March 18th 2005, to map slope failures activated or reactivated during this interval. Geographic Information Systems (GIS) were used to prepare 5 m grids of the dependent variables (absence/presence of landslide) and independent variables (lithology and 13 DEM-derivatives). Multivariate Adaptive Regression Splines (MARS) were applied to model landslide susceptibility whereas receiver operating characteristic (ROC) curves and the area under the ROC curve (*AUC*) were used to evaluate model performance. To evaluate the robustness of the whole procedure, we prepared 10 different samples of positive (landslide presence) and negative (landslide absence) cases for each area. Absences were selected through two different methods: (i) extraction from randomly distributed circles with a diameter corresponding to the mean width of the landslide source areas; and (ii) selection as randomly distributed individual grid cells. A comparison was also made between the predictive performances of models including and not including the lithology parameter.

The models trained and tested on the same area demonstrated excellent to outstanding fit ($AUC > 0.8$). On the other hand, predictive skill decreases when measured outside the calibration area, although most of the landslides occur where susceptibility is high and the overall model performance is acceptable ($AUC > 0.7$). The results also showed that the accuracy of the landslide susceptibility models is higher when lithology is included in the statistical analysis. Models whose absences were selected using random circles showed a significantly better performance when learning and validation samples were extracted from the same area; whereas, conversely, no significant difference was observed when testing the models outside the training area.

1. Introduction

Landslide susceptibility is defined as the proneness of a terrain unit to generate landslides (Brabb, 1984; Carrara et al., 1995; Guzzetti et al., 1999). A map of landslide susceptibility expresses, typically in relative terms, the spatial likelihood of landslide occurrence within a given territory. As the occurrence of slope failures may have severe economic and social consequences, landslide susceptibility maps can assist land managers and policy makers in implementing land-use strategies to reduce landslide hazard.

Landslide susceptibility may be assessed using both direct methods based on expert geomorphological analysis and indirect methods relying on deterministic or stochastic approaches. Over the last decades, the statistical approach to landslide susceptibility modeling has become very popular due to the increasing availability of low cost high-resolution data, and the development of open-source statistical software and Geographical Information Systems (GIS). This approach is based on the assumption that new landslides are more likely to occur under environmental conditions similar to those that led to past slope failures (Carrara et al., 1995; Guzzetti et al., 1999; Van Westen et al., 2005, 2008). The approach requires a landslide inventory and a set of environmental attributes related to the occurrence of slope failures. Landslide inventories are usually made by integrating field surveys with analyses of high quality aerial/satellite images. Presence or absence of landslides within a mapping unit (e.g., grid cell, slope unit, and terrain unit) represents the dependent variable, which is predicted by an ensemble of independent environmental variables. The variables are proxies of the main landslide triggering factors and are selected according to their relevance to slope stability and the quality and resolution of available data. Statistical analysis of landslide susceptibility exploits either bivariate modeling techniques (e.g., Agnesi et al., 1982; Carrara et al.,

* Corresponding author.
E-mail address: christian.conoscenti@unipa.it (C. Conoscenti).

http://dx.doi.org/10.1016/j.geomorph.2016.03.006
0169-555X/

1995; Clerici et al., 2002; Vergari et al., 2011; Rotigliano et al., 2012) or multivariate ones (e.g., Van Den Eeckhaut et al., 2006; Atkinson and Massari, 2011; Conforti et al., 2014; Cama et al., 2015; Goetz et al., 2015). Comprehensive reviews of statistical models employed in the field of landslide susceptibility modeling can be found in Aleotti and Chowdhury (1999); Guzzetti et al. (1999) and Brenning (2005).

Most of the statistical models employed to predict landslide spatial distribution are fitted to data sets with both positive (landslide presence) and negative (landslide absence) cases. Positives are often sampled from subsets of grid cells containing one to all cells within each landslide, whereas negatives are typically randomly selected as individual pixels outside the landslide areas. Then, landslide susceptibility models are calibrated and validated exploiting different samples of data, but typically extracted from the same study area, performing a random partition of positives and negatives (Chung and Fabbri, 2003). Relatively few landslide susceptibility studies have attempted a validation with independent data from areas outside those used to calibrate the models (e.g., Von Ruette et al., 2011; Pradhan et al., 2010; Costanzo et al., 2012a; Lombardo et al., 2014).

In this experiment we employed Multivariate Adaptive Regression Splines (MARS; Friedman, 1991) to model the spatial distribution of landslides that were triggered in two study areas of Sicily (Italy) by rainfall during the 2004–2005 winter season. A landslide inventory was established for both areas through the analysis of two Google Earth images, dated October 25th, 2004 and March 18th, 2005, by mapping slope failures triggered or reactivated during this time period. A limited number of studies have exploited Google Earth images to prepare landslide inventories (e.g., Costanzo et al., 2012a, 2012b; Schicker and Moon, 2012; Van Den Eeckhaut et al., 2012; Borrelli et al., 2014; Zhang et al., 2015) and, as far as we know, none produced their susceptibility models without incorporating information from field surveys and/or other sources of data. Indeed, conventional methods to prepare landslide inventory rely mainly on geomorphological field mapping and on the interpretation of stereoscopic aerial photographs (Guzzetti et al., 2012). Conversely, in this study, the 3D view provided by the Google Earth software was the only tool used for landslide detection and mapping. This allowed us to test whether effective landslide susceptibility models may be prepared without field mapping. In this experiment, we used one area to both calibrate and validate landslide susceptibility models whereas the other area was only used to assess the predictive skill of the models trained for the first area. To test the robustness of the procedure, 10 training and 10 test samples were extracted from the first area, and 10 validation subsets were identified in the second area. They were prepared by adopting different strategies to select landslide absences: (i) extraction from randomly distributed circles having a diameter corresponding to the mean width of the identified landslide source areas; and (ii) selection as randomly distributed individual grid cells. Moreover, we prepared models that both included and did not include lithology as a predictor variable. The main objectives of this experiment were to: (i) evaluate whether landslide inventories based on Google Earth images as their only data source can be used to prepare reliable landslide susceptibility models; (ii) explore how the performance of landslide susceptibility models is affected by changing the method to sample landslide absences; (iii) assess the accuracy of landslide predictions outside the area where the models were calibrated; and (iv) evaluate the importance of lithology as a predictor of landslide distribution.

2. Materials and methods

2.1. Study areas

Two study areas were selected for this experiment. Both areas are located within the catchment of the Belice River (Fig. 1), one of Sicily's main river basins. The two areas, hereafter referred to as AREA1 and AREA2 (Fig. 2), extend for 38.5 and 10.3 km^2, respectively. Their altitudes are 218–519 m a.s.l. (mean = 371.6 m and std. dev. = 68.9 m) and 317–714 m a.s.l. (481.2 and 78.8 m), respectively. The slope gradient of AREA1 (mean = 9.5° and std. dev. = 6.0°) tends to be slightly lower than that of AREA2 (10.8° and 5.6°).

The study area's climate is Mediterranean, with hot and dry summers and mild and wet winters. According to the rainfall data from the meteorological station in Corleone (588 m a.s.l.), the average annual rainfall is 643.3 mm. Precipitation occurs mainly during the autumn–winter semester, with peaks in December (91.3 mm) and January (82.5 mm) (Fig. 3).

The two study areas are mainly characterized by hilly landscapes, where slope and channel processes prevail. However, despite the very small distance separating them (around 3 km), their geological and geomorphological settings are different. AREA1 is mainly underlain by clays of the Late Miocene Terravecchia Formation (31% of the total extent) and by marls and sandstones of the Late Pliocene–Early Pleistocene Marnoso-Arenacea Formation (26%) (Table 1 and Fig. 4). The former lithology is dominant in the northern and western parts, which are characterized by gentle slopes, drained by a relatively wide and shallow valley. The latter prevails in the SW sector, where, due to the outcropping of harder rocks, the land surface is more rugged, with deeper valleys and steeper slopes. AREA2 corresponds to the lowest part of a broad valley, running approximately E–W. This area is mainly underlain by marls of the Middle–Late Miocene Marne di San Cipirello Formation and the Late Oligocene–Early Miocene Marne di Cardellia Formation (43% of the total extent). The flanks of the valley are gentle and partially covered by ancient landslides, which extend for 28% of AREA2.

Intense water erosion and gravitational processes affect both study areas. Landslides generally consist of earth-flows triggered by rainfall

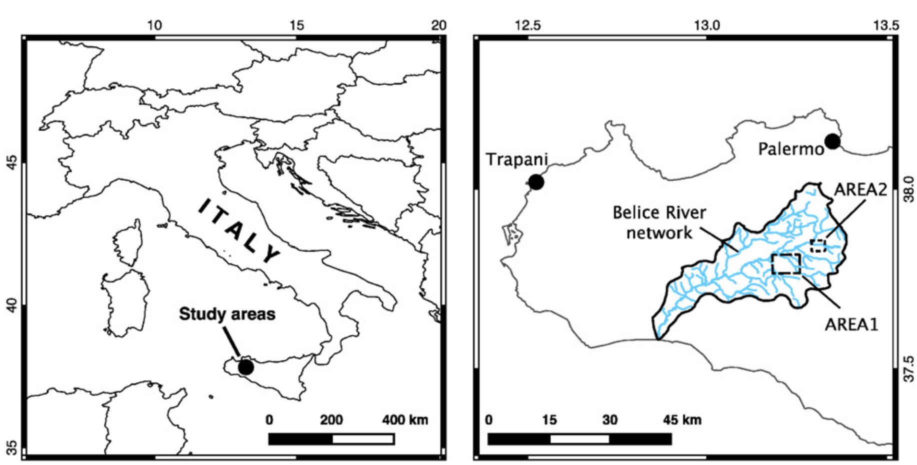

Fig. 1. Location of the study areas.