Environmental Modelling & Software 84 (2016) 467–481

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping

E. Arnone [a, *], A. Francipane [a], A. Scarbaci [b], C. Puglisi [c], L.V. Noto [a]

[a] Dipartimento di Ingegneria Civile, Ambientale, Aerospaziale, dei Materiali, Università degli Studi di Palermo, Palermo, Italy
[b] Hydro-Logic Services, Reading, Berkshire, UK
[c] Agenzia Nazionale Italiana per le Nuove Tecnologie, l'Energia e lo Sviluppo Economico Sostenibile, Rome, Italy

ARTICLE INFO

Article history:
Received 11 March 2016
Received in revised form
25 July 2016
Accepted 26 July 2016

Keywords:
Grid-cell size
Vector-to-raster conversion
Resampling
Landslide susceptibility mapping
Artificial neural network

ABSTRACT

The choice of the proper resolution in landslide susceptibility mapping is a worth considering issue. If, on the one hand, a coarse spatial resolution may describe the terrain morphologic properties with low accuracy, on the other hand, at very fine resolutions, some of the DEM-derived morphometric factors may hold an excess of details. Moreover, the landslide inventory maps are represented throughout geospatial vector data structure, therefore a conversion procedure vector-to-raster is required.

This work investigates the effects of raster resolution on the susceptibility mapping in conjunction with the use of different algorithms of vector-raster conversion. The Artificial Neural Network technique is used to carry out such analyses on two Sicilian basins. Seven resolutions and three conversion algorithms are investigated. Results indicate that the finest resolutions do not lead to the highest model performances, whereas the algorithm of conversion data may significantly affect the ANN training procedure at coarse resolutions.

1. Introduction

Landslides occur systematically all over the world, thus requiring landslide hazard researches to support an appropriate landuse planning. From this perspective, the recognition of landslide-prone terrains is traditionally considered one of the most useful approach (Hansen, 1984), which leads to the definition of landslide susceptibility maps, i.e., the spatial distribution of probability of landslide occurrence over a region. The assessment of landslide susceptibility is based on empirical or modeled relationships between landslide events occurred in the past and landslide inducing (or triggering) factors (Varnes and IAEG, 1984) of the region of interest, thus resulting into a typical spatial correlation analysis.

The scientific literature offers various methods for the landslide susceptibility mapping (Francipane et al., 2014; Arnone et al., 2014). These methods can be generally classified as: heuristic, mainly based on the opinion of geomorphologic experts (e.g., Francipane et al., 2014; Puglisi et al., 2013); bivariate statistical, mainly based on the frequency ratio approach (e.g., Lee and Pradhan, 2007; Lepore et al., 2012); multivariate statistical, mainly based on the logistic regression (e.g., Lee et al., 2004; Lee and Pradhan, 2007; Lepore et al., 2012; Francipane et al., 2014); and multivariate data-driven, mainly based on the use of Artificial Neural Network, ANN (e.g., Francipane et al., 2014; Arnone et al., 2014; Melchiorre et al., 2008). A comparison of these methods is provided by Francipane et al. (2014), who demonstrated the strength of the statistical and data-driven methods as compared to the heuristic methods.

The application of such methods requires the use of spatial distributed data generally represented by means of a raster structure. Specifically, all the input data (i.e., landslide inducing factors) and historical landslide distributions must be consistent in terms of spatial representation and resolution. The inducing factors refer commonly to continuous (e.g., elevation, slope, distance from river network or roads, etc.) or categorical (e.g., landuse, geology, soil type, etc.) variables and are described as raster data or polygons in vector data structure, respectively; events of landslide inventory, instead, are generally represented by points (landslide locations) or polygons (landslide sources and propagation area) in a vector data structure, e.g. Yilmaz (2010).

In order to make all data consistent, a sound preprocessing of data is, then, generally required, which is commonly carried out throughout spatial analysis techniques. The main issues of the preprocessing phase are i) the choice of the proper algorithm

* Corresponding author.
E-mail address: elisa.arnone@unipa.it (E. Arnone).

http://dx.doi.org/10.1016/j.envsoft.2016.07.016
1364-8152/

required to convert vector data (polygons or point features) into a raster structure and ii) the choice of the proper spatial resolution for the data.

In the case of continuous variables (e.g., elevation), a coarser spatial resolution describes the terrain less accurately, since the information of a portion of terrain identified as a pixel are synthesized with a value. This lack of accuracy is propagated on the morphological secondary attributes extracted from the DEM (Digital Elevation Model), such as the slope, the curvature, the aspect, etc. (Wu et al., 2007; Vaze et al., 2010; Yang et al., 2014). As an example, it has been demonstrated that the slope frequency distribution is very sensitive to the DEM resolution (e.g., Chang and Tsai, 1991; Kienzle, 2004; Claessens et al., 2005; Grohmann, 2015). In other words, uncertainties due to the raster structure (Pogson and Smith, 2015) of the inducing factors increase as the resolution used for the description of the variables decreases, affecting somehow the reliability of the derived susceptibility maps (Guzzetti et al., 1999; Lee et al., 2004, 2010; Yuan et al., 2008; Cama et al., 2016).

Across the studies available in literature, a large variety of resolutions, also significantly different, have been used for the development of various maps. However, not always the choice of the resolution is justified on the basis of objective criteria. Guzzetti et al. (1999) suggested the use of multiple resolutions and to opt for that which provides the best performances in terms of reliability. As demonstrated by Li and Zhou (2003), who compare the susceptibility maps obtained using data having various resolutions (from 2 m to 300 m), the best performances cannot be always attributed to the higher resolutions; in this case, the authors suggest an intermediate resolution as best option. Lee et al. (2004) considered data at 5, 10, 30, 100, and 200 m resolution, demonstrating that the higher resolutions (i.e., 5, 10, and 30 m) lead to very similar susceptibility maps. Interesting insights are given by Yuan et al. (2008), who concluded that the optimal resolution depends on the size of the study area and on the average landslide scale. In particular, in their experiment conducted on eleven different resolutions (from 5 m to 190 m), the best performance was obtained for the 90 m resolution and the worst one for the 150 m resolution; moreover, they found out that the resolution of 90 m was very close to the average area of the landslides occurred in the study area, suggesting the presence of a correlation between the best resolution and the average landslide area. Similar outcomes were found out by Lee et al. (2010), who derived landslide susceptibility maps for 12 sub-watersheds of the Chi-Shan River, southern Taiwan, at different resolutions. In this case, the pixel size derives from the spatial resolution of images from different existing sensors used to identify the landslide areas at fine resolutions (i.e., 2.5, 8, and 20 m); they, then, obtained spatial layers at coarser spatial resolutions (i.e., 10, 20, 40, 60, 80, and 100 m) by means of resampling methods. If, on the one hand, they found out that greater pixel sizes leaded to a rougher landslide map and larger landslide areas, on the other hand, they found that also the use of finer resolutions can lead to wrong results, being the optimal resolution depending on the real landslide dimensions. Recently, Cama et al. (2016) evaluated the relationship between the cell size and the model performances for susceptibility mapping of debris-flow by means of the binary logistic regression (Hosmer and Lemeshow, 2000). Specifically, they focused on the effect of landslide inducing factors on the global model performances by investigating four different resolutions (from 2 to 32 m). They found out small differences in the global accuracy and precision of models, whereas effects were detected in the role of each inducing factor as the resolution changed.

Other works explored the effect of resolution on landslide susceptibility modeling through the use of physically-based hydrological-stability models (Montgomery and Dietrich, 1994; Rosso et al., 2006; Arnone et al., 2011; Lepore et al., 2013; Capparelli and Versace, 2014; Arnone et al., 2016). By using such approaches, the cell size plays an important role mainly in the description of the terrain-based processes and the DEM derived information layers. Among these works, Claessens et al. (2005) carried out an investigation on the resolution effects using the landscape evolution model LAPSUS (Claessens et al., 2007); they concluded that a perfect resolution for a DEM cannot exist, because no resolution can generally represent perfectly the size of all landslides located throughout the analyzed area. A terrain stability model was also used by Tarolli and Tarboton (2006), who estimated the most likely initiation points for landslides from potential instability by using the Stability INdex MAPping (SINMAP, Pack et al., 1998) model and by comparing five different raster resolutions (from 2 to 50 m). They concluded that DEMs at resolutions larger than 10 m lead to a loss of resolution that degrades the results, while for DEM smaller than 10 m, the physical processes responsible for triggering landslides are obscured by the smaller scale terrain variability.

In all the above discussed works, landslide locations and areas are represented as point and polygon features, respectively, requiring a procedure to convert data from vector to raster. This aspect may considerably affect the outcome of the susceptibility analysis, even more significantly than the proper resolution that, in turn, may be related to the proper representation of the landslide polygons. Nonetheless, while the choice of the proper spatial resolution has received considerable attention, the issue of the proper representation of landslide locations and the choice of the best method to convert data from a vector to a raster structure has not been widely discussed yet.

In this work we investigated the effect of the raster resolution on susceptibility mapping in conjunction with the effect of the algorithm used to transform landslide polygons into a raster structure. We selected a well-known case study in northeastern Sicily, Italy, where, in 2009, an extreme rainfall event triggered a number of landslides on the Briga and Giampilieri basins that caused 37 casualties. Specifically, the landslide inventory consists of more than one thousands polygons, which delineate the landslide area. The inventory map was realized by the UTPRA-PREV (Technical Unit for Environmental Characterization, Prevention and Remediation – Natural risks prevention and effect mitigation) of ENEA (Italian National Agency for New Technologies, Energy, and Sustainable Economic Development) through a detailed geomorphological and morphometric field survey and an aerial photos analysis (Puglisi et al., 2013). Most of the phenomena were classified as debris and mud flows; for each event, both the landslide locations, as point information, and the landslide sources and propagation area, as polygonal information, are available. The susceptibility mapping has been conducted by applying the Artificial Neural Network technique (Arnone et al., 2014; Pumo et al., 2015) on different resolutions. Specifically, starting from the 2 m available DEM, the 10, 20, 30, 40, and 50 m DEMs have been derived, together with the associated topographic-derived variables; concurrently, three algorithms to convert vector into raster (available in ESRI GIS software) have been used and then compared.

2. Methods

2.1. Artificial Neural Network – ANN

The use of Artificial Neural Networks (ANNs) for the landslide susceptibility mapping has become very popular in literature, due especially to their efficiency in pattern recognition and multivariate analyses (e.g., Pradhan and Lee, 2010; Friedel, 2011). An ANN