Geoderma 330 (2018) 65–78

Contents lists available at ScienceDirect

Geoderma

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

Comparison of differences in resolution and sources of controlling factors for gully erosion susceptibility mapping

Younes Garosi[a], Mohsen Sheklabadi[a,*], Hamid Reza Pourghasemi[b], Ali Asghar Besalatpour[c], Christian Conoscenti[d], Kristof Van Oost[e,f]

[a] Department of Soil Science, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran
[b] Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran
[c] inter 3 GmbH - Institut for Ressources Management, Otto-Suhr-Allee 59, 10585 Berlin, Germany
[d] Department of Earth and Sea Sciences (DISTEM), University of Palermo, Via Archirafi 22, 90123 Palermo, Italy
[e] Fonds de la recherche Scientifique, FNRS, Rue d'Egmont 5, 1000 Brussels, Belgium
[f] TECLIM – Georges Lemaître Centre for Earth and Climate Research, Université catholique de Louvain, BE-1348 Louvain-la-Neuve, Belgium

ARTICLE INFO

Handling Editor: A.B. McBratney

Keywords:
Gully erosion
Machine learning model
Remote sensing data
Robustness
Topographic attribute

ABSTRACT

Gully erosion has been identified as an important soil degradation process and sediment source, especially in arid and semiarid areas. Thus, it is useful to identify the spatial occurrence of this form of water erosion in the landscape and the most vulnerable areas. In this study, we explored the effects of different pixel sizes on some controlling factors extracted from a digital elevation model and remote sensing data when producing a gully erosion susceptibility map (GESM) of Ekbatan Dam Basin, Hamadan, Iran. An inventory map of the gully landforms was prepared based on global positioning system routes of the gullies, extensive field surveys, and visual interpretations of satellite images obtained from Google Earth. Five data sets with pixel sizes ranging from 2 to 30 m were obtained using topographic attributes and remote sensing data comprising the elevation, slope degree, slope aspect, catchment area, plan curvature, profile curvature, stream power index, topographic position index, topographic wetness index, land use, and normalized difference vegetation index, which can affect the distribution of gully erosion. For each data set, 70% and 30% of the data were selected randomly for calibrating and validating the models, respectively. The statistical relationships between the occurrence of gully erosion and controlling factors were calculated using four machine-learning models, i.e., generalized linear model, boosted regression tree (BRT), multivariate adaptive regression spline, and artificial neural network (ANN). Statistical tests comprising the kappa coefficient and the area under the receiver operating characteristic curve (AUC) were calculated for both the calibration and validation data sets to estimate the optimal pixel size. The results showed that among the data sets with different pixel sizes, the optimal pixel size was 10 m for each model. In addition, the capacity of the four techniques for modeling gully erosion occurrence was quite stable when the calibration and validation samples were changed in the data set. Finally, based on three changes of the calibration and validation data sets with a pixel size of 10 m, the BRT and ANN models obtained outstanding performance (AUC $>$ 0.9), where they had the highest goodness-of-fit and predictive power, and thus the greatest robustness to changes in the calibration/validation data (i.e., lowest sensitivity to altering calibration/validation data). Our results demonstrate the importance of selecting a suitable pixel size when producing a GESM for soil and water management practices.

1. Introduction

Gully erosion has been identified as one of the most important causes of soil erosion and degradation in western Iran (Rahmati et al., 2016), as well as being a major source of the sediment delivered to streams (Poesen et al., 2003). Identifying the distribution of areas affected by gully erosion and vulnerable areas is useful for land use planning, conservation practices, or mitigating soil erosion (Conforti et al., 2011; Conoscenti et al., 2014; Pourghasemi et al., 2017; Rahmati et al., 2017). Thus, various machine learning models including bivariate and multivariate statistical methods based on *geographic information system* (GIS) data have been used for assessing the

* Corresponding author.
E-mail addresses: sheklabadi@basu.ac.ir (M. Sheklabadi), hr.pourghasemi@shirazu.ac.ir (H.R. Pourghasemi), christian.conoscenti@unipa.it (C. Conoscenti), kristof.vanoost@uclouvain.be (K. Van Oost).

https://doi.org/10.1016/j.geoderma.2018.05.027
Received 23 December 2017; Received in revised form 16 May 2018; Accepted 24 May 2018

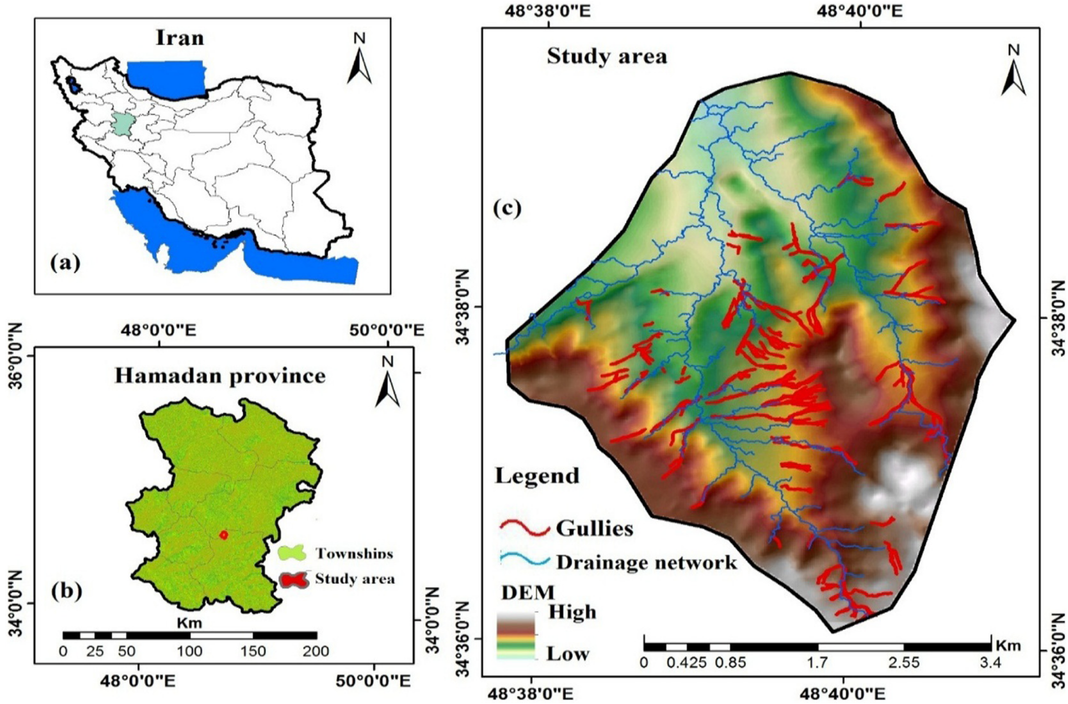

Fig. 1. Location of the study area in Hamedan Province, Iran.

susceptibility to gully erosion, such as logistic regression (Akgün and Türk, 2011; Conoscenti et al., 2014; Lucà et al., 2011), classification and regression tree (CART; Geissen et al., 2007; Gómez-Gutiérrez et al., 2009; Märker et al., 2011), random forest (Eustace et al., 2011; Rahmati et al., 2017), multivariate adaptive regression spline (MARS; Gómez-Gutiérrez et al., 2009, 2015), conditional analysis (Conoscenti et al., 2008; Magliulo, 2012; Conoscenti et al., 2013), information value (Conforti et al., 2011; Lucà et al., 2011), weights-of-evidence (Rahmati et al., 2016; Zabihi et al., 2018), frequency ratio (Rahmati et al., 2016; Zabihi et al., 2018), maximum entropy (Pourghasemi et al., 2017), support vector machine (Pourghasemi et al., 2017; Rahmati et al., 2017), and artificial neural network (ANN; Pourghasemi et al., 2017; Rahmati et al., 2017) techniques. Factors such as topographic attributes, lithology, soil properties, climate, and land use can affect the distribution of gully erosion, and thus they are often used as independent variables in these statistical models (Gómez-Gutiérrez et al., 2015).

Many studies have investigated the susceptibility to gully erosion but there is still some dispute over the most appropriate and optimal pixel size to use for controlling factors when identifying areas that are susceptible to gully erosion. Indeed, there is uncertainty about how to conduct gully erosion modeling with different pixel sizes for the controlling factors. Thus, Lucà et al. (2011) considered and compared bivariate and multivariate models of gully erosion susceptibility in Northern Calabria, Italy, and they showed that different pixel sizes affected the accuracy of the gully erosion susceptibility map (GESM), where it was dependent on both the accuracy of the controlling factors and the scale of the gullied areas. They concluded that the likelihood of classifying an area as stable or according to a low class in susceptibility mapping increased with the pixel size in the digital elevation model (DEM) employed. However, Gómez-Gutiérrez et al. (2015) noted that the availability of controlling factors at fine resolution (or a small scale) is often low, especially in underdeveloped countries or poor areas. They also found that in contrast to mapping or monitoring approaches, controlling factors with high spatial resolution did not always provide the best information for multivariate modeling purposes, whereas the relationship between the resolution of topographic attributes and the size of landforms (especially cross-sections) are important. Zhang et al. (2008) reported that the topographic and hydrological attributes extracted from DEMs with various resolutions had different properties, which could affect the results predicted by soil erosion models. Thus, selecting a suitable or appropriate resolution is very important for DEMs and other remote sensing data when studying the spatial patterns of soil erosion processes (Lu et al., 2017). In addition, the effects of different statistical techniques, mapping units, and spatial scales on the maps should be considered carefully in order to select the most suitable approach for each study area.

Statistical methods, especially multivariate techniques, are clearly suitable for obtaining quantitative estimates of the locations of future events such as landslides and gully erosion because susceptibility evaluations should be as objective as possible (Lucà et al., 2011). Among these multivariate statistical models, ANN, boosted regression tree (BRT), MARS, and generalized linear model (GLM) techniques have contributed significantly to the field of susceptibility mapping, especially studies of landslides and gully erosion (Vorpahl et al., 2012; Conforti et al., 2014; Youssef et al., 2016; Gómez-Gutiérrez et al., 2015; Pourghasemi and Rossi, 2017; Pourghasemi et al., 2017; Rahmati et al., 2017). The main advantages of these machine learning models are: (a) the dependent variable is dichotomous (i.e., 0 or 1) and the predicted values can be expressed as a probability; (b) they can use different types of independent variables such as binary categorical, ordinal, and continuous; (c) there is no requirement to define preliminary assumptions before using them; and (d) the independent variables do not necessarily need to have normal distributions when used in these models (Lucà et al., 2011; Rahmati et al., 2017).

The objectives of the present study were: (1) to obtain a series of data sets with different resolutions in order to produce a GESM using DEMs and remote sensing data from different sources; (2) to determine the statistical relationships between gully occurrence and the controlling factors for each data set using machine learning models comprising GLM, BRT, MARS, and ANN; (3) to select the best data set for each model based on goodness-of-fit and predictive performance analyses; and (4) to draw a GESM for each data set with the optimal pixel size selected by different models.