Catena 152 (2017) 144–162

Contents lists available at ScienceDirect

Catena

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

Landslide susceptibility assessment using maximum entropy model with two different data sampling methods

Aiding Kornejady [a,*], Majid Ownegh [b], Abdolreza Bahremand [a]

[a] Department of watershed management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran
[b] Department of Arid Zone Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

ARTICLE INFO

Article history:
Received 22 June 2016
Received in revised form 4 January 2017
Accepted 7 January 2017
Available online 19 January 2017

Keywords:
Random sampling
Mahalanobis distance
Predictive power
GIS
Ziarat watershed

ABSTRACT

The aim of the current study is to map landslide susceptibility over the Ziarat watershed in the Golestan Province, Iran, using Maximum Entropy (ME), as a machine learning model, with two sampling strategies: Mahalanobis distance (MEMD) and random sampling (MERS). To this aim, a total of 92 landslides in the watershed were recorded as point features using a GPS (Global Positioning System) device, along with several field surveys and available local data. By reviewing landslide-related studies and using principal component analysis, 12 landslide-controlling factors were chosen namely altitude, slope percent, slope aspect, lithological formations, proximity (to faults, streams, and roads), land use/cover, precipitation, plan and profile curvature and the state-of-the-art topo-hydrological factor known as height above the nearest drainage (HAND). Two sampling methods were used to divide landslides into two sets of training (70%) and test (30%). The Area under the success rate curve (AUSRC) and the area under the prediction rate curve (AUPRC) were used to evaluate the results of the MEMD and MERS. The results showed that both MEMD and MERS strategies with the respective AUSRC values of 0.884 and 0.878, have good performance in modelling the landslide susceptibility in the study area. However, AUPRC test showed slightly different results in which MEMD with the value of 0.906 showed excellent predictive power in comparison with the MERS with the AUPRC value of 0.846. The higher AUPRC value in relation to AUSRC indicated the MEMD as the premier model in the current study. According to the MEMD, three landslide controlling factors including lithological formations, proximity to roads and precipitation with the respective contribution percentages of 25.1%, 23.3%, and 19.1%, contained more information in relation to the rest. Moreover, according to one-by-one factor removal test, lithological formations and proximity to faults were identified to have a unique information compared to the rest. According to the MEMD, about 13.8% of the study area is located within high to very high susceptibility classes which can be matter of great interest to decision makers and the local authorities for formulating land use planning strategies and implementing pragmatic measures.

1. Introduction

Among the 40 different types of natural disasters experienced in different parts of the world, some 31 types have been identified in Iran including frequent serious earthquakes, floods, droughts, landslides, desertification, deforestation, storms and the like which are recorded due to their multiplicity, diversity, frequency, intensity, and environmental disturbance (UNISDR, 2005)[1]. According to global estimates, Iran is among the top ten countries facing natural disasters with an average of four events per year (UNISDR, 2005). Landslides are acute natural disasters like earthquakes, volcanic eruptions, and floods. Every year, they cause considerable losses to the landslide-prone areas. Over a few years or decades, landslides exhibit multiple metamorphic symptoms prior to the occurrence, rendering them as comparatively more manageable and predictable phenomena (Roy and Chacko, 1986). Thus, susceptibility assessment in basins and watersheds as a preliminary platform is an important task for future management plans.

Geologists, engineers and other specialists in various fields provide specific, or often different, definitions of landslide. The difference in definitions reflects the existence of different approaches to this phenomenon. Varnes' definition of landslide is used in the present study (Varnes and Radbruch-Hall, 1976). He defined a landslide as all downward mass movement on slopes, mainly soil, blocks of solid rock or artificial fills in response to the pull of gravity. The subject of the landslide has been constantly faced with progresses and challenges and various papers and studies have been conducted on landslide susceptibility zonation. In this research, a systematic multi-disciplinary approach has been taken to map and to analyze landslide susceptibility. To examine the rate of landslide occurrence in response to slope instability, tens of numerical

* Corresponding author.
E-mail addresses: aidin.kornejady@gmail.com (A. Kornejady), Mownegh@yahoo.com (M. Ownegh), bahremand@gau.ac.ir (A. Bahremand).

[1] United Nations Office for Disaster Risk Reduction

http://dx.doi.org/10.1016/j.catena.2017.01.010
0341-8162/

models with different parameters, weights, rates, computational algorithm, and scales have been developed, calibrated, and modified in various conditions (Ownegh, 2002, 2004). Van Westen et al. (2006) classified susceptibility zonation methods into four groups including: 1) inventory-based approaches which portray spatial and temporal patterns of landslide distribution, type of movement, rate of movement and type of displaced material through field survey mapping, historical records, satellite images and aerial photo interpretation, forming the basis for other landslide susceptibility assessment methods, 2) heuristic approaches which estimate landslide potential from expert knowledge and opinions such as bureau of Indian standards' (BIS) landslide hazard evaluation factor (LHEF), 3) statistical approaches which minimize subjectivity in weight assignment procedure and produce more objective and reproducible results including multivariate statistical approaches such as logistic regression model or bivariate approaches such as Landslide Nominal Risk Factor (LNRF), information value model, Weights of Evidence model (WofE), Frequency Ratio approach (FR), Fuzzy logic method, Shannon's entropy, Artificial Neural Network method (ANN) and Certainty Factor (CF), and 4) deterministic approaches which incorporate slope stability analysis, commonly used in site-specific models such as Factor of Safety (FS).

Maximum entropy is a statistical-probabilistic machine learning technique which has been applied to different fields such as ecology and environmental sciences especially prediction of species distribution (Phillips et al., 2004; Thuiller et al., 2005; Pueyo et al., 2007; Pearson et al., 2007; Townsend Peterson et al., 2007; Wollan et al., 2008; Kumar and Stohlgren, 2009; Williams, 2010; Elith et al., 2011; Warren and Seifert, 2011; Harte, 2011; Bajat et al., 2011; Huset, 2013; Merow et al., 2013; Yackulic et al., 2013; Yang et al., 2013; Fourcade et al., 2014; Muscarella et al., 2014; Friggens et al., 2014; Ariyanto, 2015; González-Irusta et al., 2015; Shen et al., 2015; Cao et al., 2016; Mert et al., 2016; and the references therein), crop planting regionalization (Franklin, 2010; Peterson, 2011; Yu et al., 2014; De Cauwer et al., 2014; Jaime et al., 2015), conservation policies and payment for ecosystem services (PES) schemes (Viña et al., 2013), groundwater potential mapping (Rahmati et al., 2016b) and groundwater contamination potential mapping (Wahyudi et al., 2012). Recently, several researchers have applied the maximum entropy technique to the field of landslide susceptibility assessment (Liu et al., 2012; Vorpahl et al., 2012; Felicísimo et al., 2013; Park, 2015; Convertino et al., 2013; Kim et al., 2015; Davis and Blesius, 2015; Moosavi and Niazi, 2016; Hong et al., 2016; Dickson and Perry, 2016; Pradhan et al., 2016) and the earth systems' processes (Dyke and Kleidon, 2010; Luoto et al., 2010; Brunsell and Anderson, 2011; Wang and Bras, 2011; Ruddell et al., 2013; Hjort et al., 2014; Pitchford et al., 2015; Aalto, 2015). Apart from its comparatively complex math, it does not require high precision and immense surveys. This also features the capability of predicting future spatial patterns (projection fields) by including continuous data sources such as proximity, slope degree (or percent) or precipitation layers in disregard of expert-knowledge-based classification schemes. On the other hand, to evaluate the effectiveness of the results of susceptibility assessment, landslide observations are usually divided into two data sets (landslide points) for model training and testing purposes. A number of studies concerning data sampling methods could be found, mostly considering random sampling techniques and critical sub-basin assessment approaches for result verification. As a remarkable example, Tsangaratos and Benardos (2014) separated training from the test data using Mahalanobis distance technique and introduced it as a suitable data enrichment method.

Golestan Province, due to its semi-temperate climate and the availability of geological prerequisites for the occurrence of mass movements in many different direct and indirect ways, is known as one of the critical landslide-prone provinces of Iran. This study is unique for the adoption of the maximum entropy method by two sampling techniques, called random sampling and Mahalanobis distance method, to perform an assessment of the susceptibility of the Ziarat watershed to landslide occurrence. Inclusion of the landslide susceptibility factors (either triggering or predisposing) in this research draws heavily on an extensive literature review, expert knowledge, and several field reconnaissance. Other features of the current research is the adoption of the new topo-hydrological factor "HAND" which have several advantages to the traditional DEM layers in terms of involving several attributes along with the terrain information. Besides its applications in hydrology (Rennó et al., 2008; Savenije, 2010; Nobre et al., 2011; Gharari et al., 2011; Cuartas et al., 2012), a number of innovative applications have been found in other environmental disciplines. Besides elevation, HAND identifies different hydrologic response zones and runoff generation mechanisms; making it suitable to be used along with other water-related factors such as proximity to streams. The possibility exists to replace the old-fashioned topo-hydrological factors such as topographic wetness index (TWI) with the HAND. The final stage of the study comprised the evaluation and confirmation of model results by means of the success and prediction rate curves. The best model was then used to extract more results on factors' importance and spatial response curves.

2. Study area

As a mountainous area in Golestan province, the Ziarat watershed, drains a 91.48 km^2 basin. This watershed lies within 4,055,495–4,071,815 m UTM zone 40 latitudes and 267,542–278,252 m longitudes, with the minimum elevation of 558 m and a maximum elevation of 3027 m above sea level. The study area is located in the southern part of the city of Gorgan and forms one of the tributaries of the Gharesoo River (Fig. 1). The Ziarat watershed contributes to the supply of 30% of the Gorgan city's domestic water demand. Yet, landslides have interfered with the development activities in the watershed, mainly in the form of lateral slips along the watercourses that has resulted in the reduction of the effective lifespan of the reservoirs. Despite the possible environmental risks, this watershed is becoming a tourism hub by the development of settlements and tourist facilities. In spite of the small area of the landslides compared to the total study area (0.3%), several symptoms exists in the region testifying to the active nature of landslides. For instance, there are numerous reports on the emergence of cracks in houses and occurrence of heavy rock falls in a part of the Ziarat - Gorgan road (Fig. 3F). The abovementioned problems justify the need to conduct such studies. The main channel of the Ziarat watershed, flowing in SE-N direction, drains most of the precipitations. Average annual precipitation amounts to 550 mm. Maximum and minimum precipitation occurs in December (49 mm) and August (18.6 mm), respectively. The average slope is about 48%. The watershed is devoted to various land uses/covers including forests, rangelands, farmlands and residential areas. Covering almost 36% of the study area, Gorgan Schists and Shemshak formation, with high soil erodibility, have the highest sediment production rate in the Toolbone subwatershed. This watershed has experienced a remarkable change in land-use/cover over the recent years in the form of the expansions of settlement areas.

The Ziarat watershed consists of young deposits placed on the oldest rocks with an unconformable contacts. The study area includes several formations with the oldest one related to Post Ordovician to pre-lower Carboniferous period composed of intrusive rocks in the northern areas and the recent formation composed of Quaternary alluvial deposits. The general strike of the formations has expanded in a northeast-southwest direction. A more detailed description of these formations is provided in Section 3.2.8. With regards to the erosion and weathering, the rock outcrops of the study area are classified into four metamorphic, calcareous, clastic and alluvial groups. The Calcareous unit covers the majority of the area (72.39%) while the metamorphic unit expands less than other groups (1.41%). Calcareous formations include the Khosh-Yeilagh (Dkh), Mobarak (Cml), Dorud (Pdl), Ruteh (Pr), Lar (Jl1), and the Upper Cretaceous limestone. Changes in temperature and high humidity in the metamorphic rocks are