Catena 157 (2017) 213–226

Contents lists available at ScienceDirect

Catena

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

A comparative study between popular statistical and machine learning methods for simulating volume of landslides

Ataollah Shirzadi[a], Himan Shahabi[b,*], Kamran Chapi[a], Dieu Tien Bui[c], Binh Thai Pham[d], Kaka Shahedi[e], Baharin Bin Ahmad[f]

[a] Department of Rangeland and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran
[b] Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran
[c] Geographic Information System Group, Department of Business Administration and Computer Science, University College of Southeast Norway, Hallvard Eikas Plass, N-3800 Bøi, Telemark, Norway
[d] Department of Geotechnical Engineering, University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Ha Noi, Viet Nam
[e] Department of Watershed Sciences Engineering, Faculty of Natural Resources, University of Agricultural Science and Natural Resources of Sari, Sari, Iran
[f] Department of Geoinformation, Faculty of Geoinformation and Real Estate, Universiti Teknologi Malaysia (UTM), Malaysia

ARTICLE INFO

Keywords:
Landslide
Simple statistical models
Machine learning algorithms
ANFIS
Kurdistan province
Iran

ABSTRACT

This study attempts to compare popular statistical methods (linear, logarithmic, quadratic, power and exponential functions) with machine learning methods (multi-layer perceptron (MLP), radial base function (RBF), adaptive neural-based fuzzy inference system (ANFIS) and support vector machine (SVM)) for simulating the volume of landslides based on their surface area (VL ~ AL) in the Kurdistan province, Iran. Performances of the models were validated using some commonly error functions including the Adjusted R^2, F-test and AIC (Akaike Information Criteria). The results showed that the power model demonstrates the best performance compared to other statistical methods whereas the ANFIS model outperforms other machine learning approaches. Furthermore, the comparative results showed that machine learning methods indicate better performances than simple statistical methods for simulating the volume of landslides in the study area. In practice, the outputs of this research can help managers and investigators decrease the cost of field surveys and measurements of volumes of landslides in landslide hazard management projects.

1. Introduction

Landslides are a complex natural phenomenon causing huge losses of human lives and properties in many countries over the world (Lee et al., 2006; Shahabi et al., 2013). They play a major role in the evolution of landforms. Landslides have different types such as rock fall, rock avalanche, soil slip, mud flow, etc. Despite landslides occur frequently and widespread, the statistics of landslide sizes are not easily determined (Chau et al., 2004; Malamud et al., 2004; Shahabi and Hashim, 2015; Van Den Eeckhaut et al., 2007; Hungr et al., 2014). It is obvious that through interpretation of aerial photographs and satellite images, assessing surface area of landslide (AL) and number of landslides (NL) are possible, however determination of the volume of landslides (VL) is still a challenge (Hungr et al., 1999; Issler et al., 2005; Dai and Lee, 2001; Fujii, 1969). Recently, several researchers have used the statistical relationships between landslide occurrence and rainfall to estimate volumes of landslides (Guthrie and Evans, 2004; Hovius et al., 1997; Imaizumi et al., 2008; Korup, 2005; Shahabi et al., 2014). Literature review shows that the statistical relationships between the NL and VL (Brunetti et al., 2009; Guzzetti et al., 2004; Othman and Gloaguen, 2013; Simonett, 1967), and also relationship between AL and VL (Haflidason et al., 2005; Larsen and Torres-Sanchez, 1998; Rice and Foggin, 1971; Whitehouse, 1983) seem to follow a single relation and a power law.

Recently, machine learning methods have been used for various hazard assessments and geo-engineering applications (Lee, 2007; Melchiorre et al., 2008; Oh et al., 2010). For instance, They have been applied to landslide studies (e.g., Falaschi et al., 2009; Lee et al., 2006; Nefeslioglu et al., 2008; Pradhan and Lee, 2010; Zare et al., 2013), generate landslide susceptibility mapping (Alimohammadlou et al., 2014; Bui et al., 2016; Conforti et al., 2014; Goetz et al., 2015; Micheletti et al., 2014; Nourani et al., 2014; Shirzadi et al., 2017a) as well as to produce rock fall susceptibility mapping (Shirzadi et al., 2012, 2017b). Although machine learning techniques have been broadly applied to many fields of natural sciences, their applications on the relationship analysis between the geometrical features of land-

* Corresponding author.
E-mail address: h.shahabi@uok.ac.ir (H. Shahabi).

http://dx.doi.org/10.1016/j.catena.2017.05.016
Received 4 September 2016; Received in revised form 19 April 2017; Accepted 15 May 2017
Available online 25 May 2017

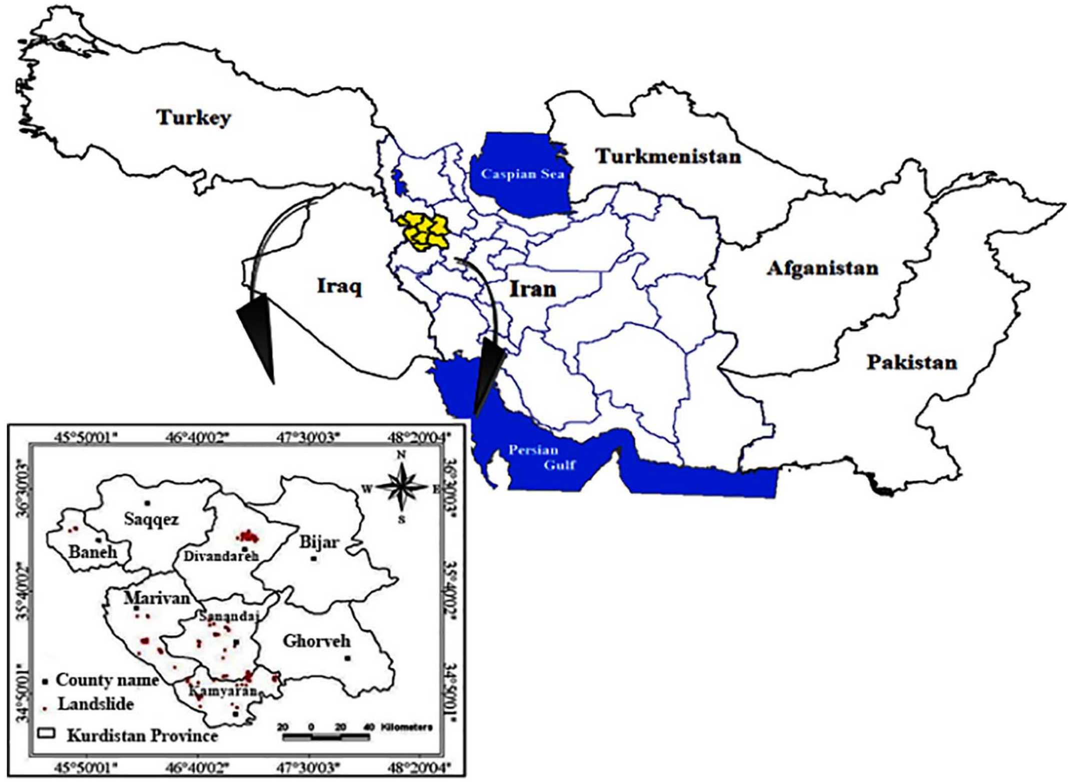

Fig. 1. Location of landslides in the some counties of Kurdistan province, Iran.

slides, for example surface area and volume of landslides, have not been reported yet. Therefore, this study aims to deal with this challenge and tends to evaluate the capability of machine learning methods for simulating the volume of landslides based on the surface area of landslides (VL ~ AL) in some counties of the Kurdistan province, Iran through comparing statistical methods (linear, logarithmic, quadratic, power and exponential functions) and machine learning methods (multi-layer perceptron neural networks (MLP), radial basis function neural networks (RBF), adaptive neural-based fuzzy inference system (ANFIS), and support vector machine (SVM).

2. Study area and data used

2.1. Description of the study area

Kurdistan province is located between longitudes of 45°33′58.51″ and 48°13′53.13″ N, and between latitudes 34°45′05.69″ and 36°27′57.56″ E. It lies in the northwest of Iran, covering an area 29,137 km^2 (Fig. 1). The elevation varies from 900 m to 3300 m a.s.l. Kurdistan province is a mountainous area with fairly high precipitation (average annual of about 500 mm) experiencing many landslides in recent years. The study area is divided into two parts namely east and west. In the east part, there are plains with low elevation and proper fertile soil. The annual mean precipitation of eastern part is about 300 mm. In the west, the topography is complex including mountains with high elevation. The annual average precipitation is about 800 mm.

The climate of Kurdistan is affected by the Mediterranean warm and humid weather which leads to heavy rainfall in the spring and snowfall in the winter. The annual mean temperature of the province is 13 °C. The number of frost days in the Kurdistan province is 92 days. Land covers include forest lands (mainly covered by forest stands of *Quercus castaneifolia*, *Quercus infectoria* and *Quercus libani*), rangelands, and farmlands.

Kurdistan province can be divided into three parts based on the geological and morphotectonic perspective. The first part is the east and southeast covered by Permian sediments (carbonate rocks and low metamorphic Schist), Jurassic (Shale and sandstone, limestone and volcanic), Cretaceous (Thin layer of limestone), and Miocene deposits (Organic limestone, marl, and evaporate sediments). The second part is located in the central and north of province where the morphology is almost uniform. The oldest rocks are metamorphic series (Gneisses) of Precambrian. The third part is the south and southwest where the oldest rocks consist of the Triassic series and a series of fine-grained gray limestone; dolomite and dolomitic mudstone with a low thickness which has been located underneath the subset of Jurassic and Cretaceous deposits.

2.2. Data acquisition and methods

The determination of AL and NL is relatively simple; whereas, determination of VL is still a challenge. Landslide inventory map for the study area was constructed such that landslide-polygons were identified by interpretation of aerial photographs and satellite images. These landslide locations were then checked by field investigation. In addition, landslides collected based on the geophysical exploration and field surveying from the Forests, Range and Watershed Management Organization of Iran (FRWMO) of Iran were also included (Table 1).

The AL and VL were checked based on the landslide inventory map and directly determined using field surveys (geophysics explores). Surface areas of landslides were recorded using GPS during field surveys and then digitized in ArcGIS 10 software by "spatial statistics tools" command. The volumes of landslides were obtained using depth of occurred landslides. The latter itself was estimated by experts. Different landslide types occurred including rotational sliding, translating sliding, complex, flow, rotational falling, and lateral spreading (based on Varnes (1978) classification) were identified during field surveys (See Table 1).

Landslide volumes ranged from $2.5 * 10^1$ m^3 to 10^7 m^3. The total number of landslides occurred in the Kurdistan province is 370, including 46, 57, 49 and 189 landslides in Sanandaj, Kamyaran,