Catena 114 (2014) 21–36

Contents lists available at ScienceDirect

Catena

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

A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping

Omar F. Althuwaynee [a], Biswajeet Pradhan [a,*], Hyuck-Jin Park [b,*], Jung Hyun Lee [b]

[a] Department of Civil Engineering, Faculty of Engineering, University Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia
[b] Department of Geoinformation Engineering, Sejong University, 98 Gunja-dong, Gwangjin-gu, Seoul 143747, South Korea

ARTICLE INFO

Article history:
Received 13 June 2013
Received in revised form 7 October 2013
Accepted 28 October 2013

Keywords:
Ensemble model
Landslide susceptibility
Remote sensing
Geographic information system (GIS)
South Korea

ABSTRACT

This study compares the landslide susceptibility maps from four application models, namely, (1) the bivariate model of the Dempster–Shafer based evidential belief function (EBF); (2) integration of the EBF in the knowledge-based analytical hierarchy process (AHP) as a pairwise comparison model processed by using all available causative factors; (3) integration of the EBF in the knowledge-based AHP as a pairwise comparison model by using high nominated causative factor weights only; and (4) integrated EBF in the logistic regression (LR) as a multivariate model by using nominated causative factor weights only. These models were tested in Pohang and Gyeongju Cities (South Korea) by using the geographic information system GIS platform. In the first step, a landslide inventory map consisting of 296 landslide locations were prepared from various data sources. Then, a total of 15 landslide causative factors (slope angle, slope aspect, curvature, surface roughness, altitude, distance from drainages, stream power index, topographic wetness index, wood age, wood diameter, wood type, forest density, soil thickness, soil texture, and soil drainage) were extracted from the database and then converted into a raster. Final susceptibility maps exhibit close results from the two models. Models 1 and 3 predicted 82.3% and 80% of testing data during the analysis, respectively. Thus, Models 1 and 3 show better performance than LR. These resultant maps can be used to extend the capability of bivariate statistical based model, by finding the relationship between each single conditioning factor and landslide locations, moreover, the proposed ensemble model can be used to show the inter-relationships importance between each conditioning factors, without the need to refer to the multivariate statistic. The research outcome may provide powerful tools for natural hazard assessment and land use planning.

1. Introduction

Natural and man-made hazards are the most threatening factors to human life and are the main reasons behind serious worldwide disasters (Akgun and Kincal, 2010; Aleotti and Chowdhury, 1999; Althuwaynee et al., 2012b; Devkota et al., 2013; Mugagga et al., 2012; Oh and Pradhan, 2011). Frequent heavy rainfalls throughout the years have produced significant slope failures in susceptible areas. In 1996, 1998, and 1999, serious damages to people and infrastructure were registered as catastrophic economical loses in South Korea (Lee et al., 2007). Landslide inventory data can be evaluated against a number of conditioning factors to identify their roles and consequences in controlling the occurrence of landslides (Broothaerts et al., 2012; Che et al., 2012; Mugagga et al., 2012; Tien Bui et al., 2012a, 2012b, 2012c; Zare et al., 2012).

Despite considerable literature on landslide hazard assessment and the development of many statistical techniques, the detection and prevention of landslides are still very much open topics (Van Westen et al., 2006). Thus, landslide researches need to focus on determining the type and quantity of triggering factors to mitigate losses and investigate the geomorphological, physical, and topographical factors that influences the occurrence of tragedies. Thereafter, recommendations and measurements should be proposed to select the best remedy. However, studies on landslides in South Korea are still lacking compared with other studies in landslide prone areas worldwide (Lee et al., 2013).

The effectiveness of landslide studies around the world is apparent from the high prediction results of landslide susceptibility maps and assessment reports from models such as the multivariate logistic regression (LR) (Ayalew and Yamagishi, 2005; Tien Bui et al., 2011; Pradhan, 2010; Pourghasemi et al., 2013a), bivariate evidential belief function (EBF) (Mohammady et al., 2012; Tien Bui et al., 2012b), and knowledge-based analytical hierarchy process (AHP) (Ghosh et al., 2011; Kayastha et al., 2012; Pourghasemi et al., 2012a). Statistical approaches and knowledge-based data analysis are used and incorporated in recent studies, as well as the geographic information systems (GIS), to reduce subjectivity (Akgun et al., 2012; Pourghasemi et al., 2012b, 2013b; Pradhan et al., 2010, 2011; Tien Bui et al., 2011, 2012a,c).

* Corresponding authors.
E-mail addresses: biswajeet24@gmail.com, biswajeet@lycos.com (B. Pradhan), hjpark@sejong.ac.kr (H.-J. Park).

0341-8162/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.catena.2013.10.011

The idea of building a predictive model by integrating multiple models (M_1 to M_n) with the aim of creating an improved model M*, has been under investigation by different researchers. Bühlmann and Yu (2003) used an ensemble of two linear regression models. Dimitriadou et al. (2003) pointed out that robustness and clustering algorithms can be positively affected by ensemble methods. These ensemble methods can effectively make use of such diversity to reduce the variance-error (Tumer and Ghosh, 2001) without increasing the bias-error. In another paper, Bartlett and Shawe-Taylor (1999) used theory of large margin classifiers coupled with ensembled technique to reduce the bias-error effectively. In a recent paper, Ghosh and Acharya (2011) explored many algorithms on consensus clustering. Rokach (2010) reviewed many existing ensemble techniques applied in various disciplines. In a very recent paper, Tehrany et al. (2013) applied ensemble rule based decision tree (DT) and bivariate and multivariate statistical models in spatial prediction of flood areas in a tropical area of Malaysia. Their result indicated that the ensemble model could provide reasonable good prediction accuracy in identifying flood susceptible areas. Similarly, Tien Bui et al. (2013) applied J48 ensemble DT and its ensemble model for landslide susceptibility mapping in a study area in Vietnam.

This study is inspired from the work conducted by Ghosh et al. (2011). This study proposes the use of EBF as a bivariate method to investigate the integration validity with AHP as a knowledge-based pairwise comparison method and LR as a multivariate method for predictive map generation (Ayalew and Yamagishi, 2005).

The EBF, together with the Dempster–Shafer theory of belief, was successfully used and validated as an effective model for landslide susceptibility map generation (Althuwaynee et al., 2012a; Park, 2010). The Dempster–Shafer theory of belief provides the framework for the estimation of EBFs, which are integrated according to Dempster (1968) rule of combination as a bivariate statistical analysis method. EBFs are generally applied in mineral potential mapping (Carranza, 2009). In this method, the ranking of influencing parameters was conducted by using landslide densities (Ayalew and Yamagishi, 2005).

The LR model is considered one of the main multivariate models and is widely employed for landslide susceptibility mapping (Bai et al., 2011; Domínguez-Cuesta et al., 2007; Pradhan, 2010). The LR method can perform variable regression even if the variable has an abnormal distribution (Menard, 2000). Furthermore, the current application procedure of LR simplifies the interpretation of results by pre-processing factors with EBF.

The AHP (Saaty, 1980; Yalcin et al., 2011) is considered as the main tool in the subjective assessment of qualitative methods. The descriptive expression of AHP mainly depends on the knowledge of experts. A pairwise comparison matrix is assigned for each parameter and subcriterion from the dual comparison results. The AHP process is based on three ideologies: decomposition, comparative judgment, and synthesis (Malczewski, 1999). The synthesis ideology compares pairs of decision factor hierarchy to assign the weights and examine the consistency ratio (CR). However, AHP is highly subjective because it includes virtual assessments that depend on the level of expert experience in parameter ranking (Kayastha et al., 2012).

Historical landslide inventory data can be evaluated against a number of conditioning factors to identify the roles and importance of these factors in the occurrence of future landslides. The methodology flow starts by classifying the inventory map into training locations (calibration) and data validation locations (testing) by using EBF (Althuwaynee et al., 2012a) to rank quantitatively the individual spatial factor classes. Thereafter, a pairwise relationship is generated between the spatial factors by integrating the EBF weights in AHP. To show the reliability of the proposed work, the resultant maps were compared by using the LR-based multivariate statistical analysis method on EBF class weights.

The methodology applications assessed the slope failures in Pohang and Gyeongju Cities (South Korea). A slight difference in results was shown between the integrated model and LR because of the difference in the nature of the statistical analysis. The possibility of adopting this theory could be appreciated in a medium-scale landslide prediction assessment.

2. Study area

The study area is located in North Gyeongsang Province, which shares a border with the Pohang basin and Gyeongju basin. The study area had an approximate area of 170 km^2 located between 129° 17′ and 129° 25′ west–east longitudes and 36° 50′ and 35° 55′ north–south latitudes (Fig. 1). The altitudes reach up to 420 m in the western parts of the region. The landslide types mainly consist of 89% translational landslide, 10% mudflow, and 1% rotational slope failure. Table 1 illustrates the main characteristics of translational landslides in the study area.

Generally, the travel distance of landslide debris in the study area is short, whereas the depth of soil properties is approximately less than 50 cm, which is considered shallow. Large amounts of soil materials have already been removed from the slope. Thus, the slopes are significantly exposed to weathering and erosion for the entire year. The debris is often deposited downward at the plains and is transported along the slope surface.

The climate characteristics in the study area are the average of the climate in the two basins: hot summer (May to late August) and cool winter (December to late February). The average annual temperature during a monsoon season (late June to early August) is 12.2 °C (54.0 °F). The average registered humidity is 68% from May to September annually. Pohang and Gyeongju exhibit highly intense precipitation that releases trapped water level in slopes, that increase the shear stress but causing a low shear strength in debris mass materials (Varnes, 1984). The study area receives the highest amount of rainfall from June to September (114 mm to 326 mm). An average precipitation of 150 mm was recorded in 2 days from 25 July 1998 to 26 July 1998. A total of 283 landslides were recorded during this time (Korea Meteorological Administration).

Geologically, the study area is composed of mudrock with abundant shale that has been deposited in shallow marine basins during the tertiary period. Slopes are vulnerable to weathering because mudrock contains abundant clays that display fragile, slaking, and swelling properties and are sensitive to extrinsic conditions such as rainfall or snow all-year around. Therefore, the geotechnical properties of mudrock, such as compression strength, can be easily modified and are susceptible to landslide. Landslides in mudrock, gneiss, and granite areas occur in a similar manner; that is, these landslides initially develop from translational slides. Thereafter, these landslides develop into flow-type landslides down the slope in the sliding stage.

3. Data

Landslides are classified by type of movement. The total inventory data showed that 296 sites represented as polygons that were mapped on a 1:5000 scale, wherein 40% of the total landslide was found on fine clay and 20% was found in loamy skeletal soil texture. Table 1 shows the main characteristics of translational landslide in the study area.

Fifteen causative factors, such as slope angle, slope aspect, curvature, surface roughness, altitude, distance from drainage, stream power index (SPI), topographic wetness index (TWI), wood age, wood diameter, wood type, forest density, soil thickness, soil texture, and soil drainage, were mapped to generate the final susceptibility maps (Fig. 2). The structural control of lineaments is not remarkable in the Pohang and Gyeongju Cities. Thus, information about lineaments such as distance from the lineament or density was not recognized as a causative factor in our study.