Geoscience Frontiers 6 (2015) 807–816

HOSTED BY
ELSEVIER

Contents lists available at ScienceDirect

China University of Geosciences (Beijing)

Geoscience Frontiers

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

Research paper

Discrete rough set analysis of two different soil-behavior-induced landslides in National Shei-Pa Park, Taiwan

Shih-Hsun Chang [a], Shiuan Wan [b,*]

[a] Information Networking and System Administration, Ling Tung University, Taiwan
[b] Information Management, Ling Tung University, Taiwan

ARTICLE INFO

Article history:
Received 8 August 2013
Received in revised form
27 November 2013
Accepted 16 December 2013
Available online 4 February 2014

Keywords:
Landslide
Data mining
Discrete rough sets
Taiwan

ABSTRACT

The governing factors that influence landslide occurrences are complicated by the different soil conditions at various sites. To resolve the problem, this study focused on spatial information technology to collect data and information on geology. GIS, remote sensing and digital elevation model (DEM) were used in combination to extract the attribute values of the surface material in the vast study area of Shei-Pa National Park, Taiwan. The factors influencing landslides were collected and quantification values computed. The major soil component of loam and gravel in the Shei-Pa area resulted in different landslide problems. The major factors were successfully extracted from the influencing factors. Finally, the discrete rough set (DRS) classifier was used as a tool to find the threshold of each attribute contributing to landslide occurrence, based upon the knowledge database. This rule-based knowledge database provides an effective and urgent system to manage landslides. NDVI (Normalized Difference Vegetation Index), VI (Vegetation Index), elevation, and distance from the road are the four major influencing factors for landslide occurrence. The landslide hazard potential diagrams (landslide susceptibility maps) were drawn and a rational accuracy rate of landslide was calculated. This study thus offers a systematic solution to the investigation of landslide disasters.

1. Introduction

Mass movements cause one to two billion dollars in financial loss each year in Taiwan. The resultant of this has been studied attempting to predict landslide occurrence at landscape scales, which in turn has led to the development of numerous stochastic, statistical and model-based simulations, with increasing emphasis on the use of GIS, in the past 10 years (Aleotti and Chowdhury, 1999; Guzzetti et al., 1999). Different modeling approaches have generally taken the form of multivariate statistical analyses of landscape characteristics associated with past landslide records (Lee and Min, 2001; Santacana et al., 2003). Parameters required for either process-based or stability models typically include data on slope, cohesion, and soil moisture or surrogate measures of those variables (Anbalagan, 1992). Slope components in GIS-based models are most often estimated directly from digital elevation models (Soeters and Van Westen, 1996; Barredo et al., 2000). Different soil types also govern the stability problem by offering different slope positions for mass movement (Rupke et al., 1988; Dai and Lee, 2002; Ma et al., 2010).

Accordingly, landslide is the major catastrophic hazard which has drawn the attraction of geoscientists, engineering professionals and decision makers (Cruden and Varnes, 1996; Akgun and Turk, 2010; Bouaziz et al., 2011). With the growing techniques of spatial data survey in geosciences, large amounts of data or information can easily be collected and monitored (Suzen and Doyuran, 2004; Prasannakumar et al., 2012). Thus, the collection of influencing variables of in-situ data becomes easier than usual. The variables of depictive data in a given zone have uncertainties which require preprocessing analysis to achieve better accuracy (Zerger, 2002; Tangestani, 2004). There are two possible techniques to handle the complicated monitoring data: dimension reduction and clustering analysis (Goktepe et al., 2005). The former reduces the dimensionality of the considered information systems (Lei et al., 2007; Lin, 2008; Wan et al., 2008, 2009). It irreversibly transforms the descriptive dataset features. The latter consists of researches developed across a wide variety of communities by

* Corresponding author. Tel.: +886 4 24629863.
E-mail address: shiuan123@mail.ltu.edu.tw (S. Wan).
Peer-review under responsibility of China University of Geosciences (Beijing).

1674-9871/$ – see front matter © 2015, China University of Geosciences (Beijing) and Peking University. Production and hosting by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gsf.2013.12.010

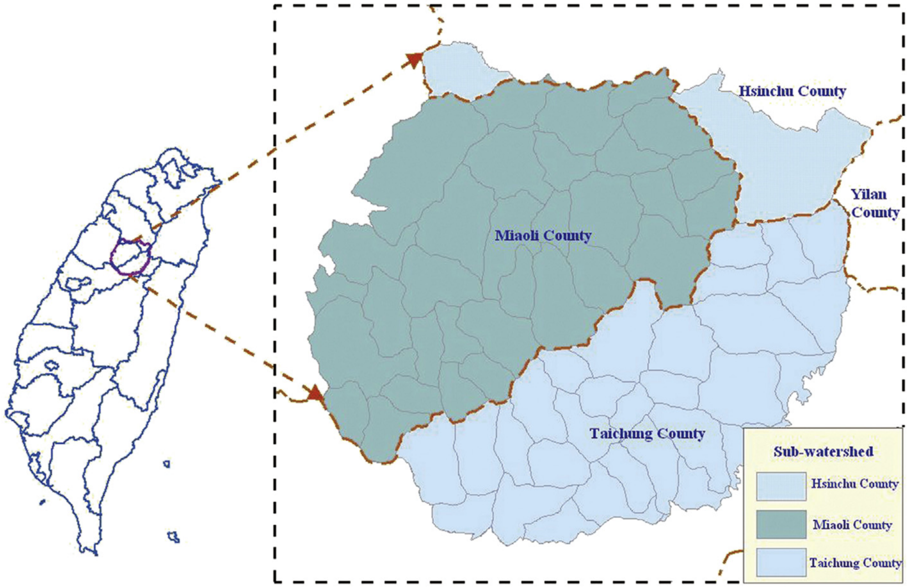

Figure 1. Shei-Pa watershed and its geographical location (Lei, 2004).

primitive exploration with little or no prior knowledge (Pradhan et al., 2010).

Several different techniques for landslide susceptibility mapping have been proposed (Wan, 2009; Pradhan et al., 2010; Yilmaz, 2010; Wan et al., 2012). Some of the researchers focused on the quantitative techniques, such as statistical methods (Lee and Sambath, 2006; Chang et al., 2010), deterministic approaches (Carrara et al., 2003) and so forth (Lee et al., 2003; Lee et al., 2004; Melchiorre, 2008; Pradhan and Lee, 2010; Pradhan et al., 2011). Generally, landslides are mostly associated with sedimentary rocks subjected to a primary driving force (rainfall or earthquake excitation) for their occurrence (Khazai and Sitar, 2003; Zhou and Li, 2011). Based on the past observations, landslide concentration values seem to diminish beyond distances of 40 and 70 km from the epicenter and the surface projection of the fault plane, respectively. To sum up, the occurrence of landslide events may be caused by triggering factors such as heavy rainfall and earthquake ground motions as well as induced factors such as fragile geology of steep slopes along fault planes (Varnes, 1978; Van Westen et al., 1997; Iverson, 2000). In recent years, data mining approaches (Lei et al., 2008; Wan et al., 2008) have offered a brand new solution to analyze landslides (Wan, 2009) and other geological phenomena (Wan et al., 2010). Our research used the discrete rough set (DRS) method (Pawlak, 1991; Nguyen and Skowron, 1995) to tackle the uncertainty problem from the materials and parameters involved in an observed landslide (Wan et al., 2012). The discrete rough set (for explanation of DRS refer to Nguyen and Nguyen, 1998) is completely different from the traditional deterministic or statistical methods that require ‘weights’ and parametric studies among independent variables. We used cross-validation methods to find the ‘best’ training datasets from the database. Thus, the best knowledge rules can be achieved and the performance of the knowledge rules also verified.

2. The development of Shei-Pa National Park database

2.1. Study area

The study area is the Shei-Pa National Park and the data were collected in the period after the occurrence of the Chi-Chi earthquake. Shei-Pa National Park (area: 75,000 ha) is situated 40–80 km from ChelungPu Fault, which transmitted tremendous energy to the central part of Taiwan (see Fig. 1). The landslide zone in our study area has relatively shallow slides on very steep slopes underlain by stiff soils and jointed rock. Utilizing DEMs, SPOT-image data, we investigated maps and GIS spatial attribute

Figure 2. Landslide spot (after Aere Typhoon, 2004; Source: Shei-Pa National Park Management Division). (a) Landslide area (gravel falls); (b) landslide area (slides for loam).