Computers & Geosciences 37 (2011) 1311–1317

Contents lists available at ScienceDirect

Computers & Geosciences

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

A rough set approach to analyze factors affecting landslide incidence

J.P. Liu [a], Z.P. Zeng [b], H.Q. Liu [a], H.B. Wang [c,d,*]

[a] Resource and Environmental Science of Wuhan University, Wuhan 430070, Hubei, People's Republic of China
[b] College of Public Administration, Huazhong University of Science and Technology, Wuhan 430074, Hubei, People's Republic of China
[c] School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, People's Republic of China
[d] Hubei Key Laboratory of Structural Engineering, Huazhong University of Science and Technology, Wuhan 430074, Hubei, People's Republic of China

ARTICLE INFO

Article history:
Received 6 May 2010
Received in revised form
11 January 2011
Accepted 19 February 2011
Available online 15 March 2011

Keywords:
Landslides
Spatial information
Data mining
Rough set theory

ABSTRACT

Landslide incidence can be affected by a variety of environmental factors. Past studies have focused on the identification of these environmental factors, but most are based on statistical analysis. In this paper, spatial information techniques were applied to a case study of landslide occurrence in China by combining remote sensing and geographical information systems with an innovative data mining approach (rough set theory) and statistical analyses. Core and reducts of data attributes were obtained by data mining based on rough set theory. Rules for the impact factors, which can contribute to landslide occurrence, were generated from the landslide knowledge database. It was found that all 11 rules can be classified as both exact and approximate rules. In terms of importance, three main rules were then extracted as the key decision-making rules for landslide predictions. Meanwhile, the relationship between landslide occurrence and environmental factors was statistically analyzed to validate the accuracy of rules extracted by the rough set-based method. It was shown that the rough set-based approach is of use in analyzing environmental factors affecting landslide occurrence, and thus facilitates the decision-making process for landslide prediction.

1. Introduction

Landslides can be triggered by earthquake ground motion, heavy rainfall, and numerous other factors. In addition, environmental conditions can also contribute to landslide occurrence, such as the slope geology and morphological parameters. To identify landslide susceptibility, past studies have focused on statistical methods, including the likelihood ratio, analytical hierarchy method, and artificial neural networks (e.g., Chung, 2006; Nefeslioglu et al., 2008; Melchiorre et al., 2008; Yilmaz, 2009; Wu and Chen, 2009). However, the disadvantage of these approaches is that the important variables were statistically identified, and/or their values and relative weightings were statistically assigned in terms of their perceived influence on the incidence of landslides. In addition, these methods incorporated all the environmental factors. In reality, only a few factors or combinations of factors had a great contribution to the incidence of landslides. It is crucial to identify these factors, so that landslide susceptibility can be accurately mapped.

Recent research progress in database technologies has created interest in knowledge discovery for landslide susceptibility prediction (Gorsevski and Jankowski, 2008; Saito et al., 2009; Wan, 2009; Wan et al., 2009). For example, landslide incidence data can be categorized and thoroughly evaluated by an entropy-based classification method to construct landslide knowledge rules. Intuitively, knowledge for landslide prediction can be perceived as a body of information, which constitutes our domain of interest. The knowledge representation system can be perceived as a data table, columns of which are labeled by attributes, and rows by landslide presence or no landslide presence. The decision problems for the selection of environmental factors can be formulated using the above decision table formalism; thus, this tool is particularly useful in landslide susceptibility prediction. In the present study, the rough set theory developed by Pawlak (1982) was considered for reduction of a decision table in such a manner that some decisions can be made with a smaller number of attributes, i.e., with a reduced number of environmental factors or combination of factors affecting the presence of landslides. In the rough set processing, a landslide susceptibility database is regarded as a decision table, which is made up of the universe or discourse, a family of equivalent relations over the universe, condition attributes (factors affecting landsliding), and decision attributes (landslide presence or no landslide presence).

The rough set processing is completely different from the traditional statistical analyses that assumed distributions in the

* Corresponding author at: School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, People's Republic of China.
E-mail addresses: huabin4631@yahoo.com.cn, huabin@mail.hust.edu.cn (H.B. Wang).

0098-3004/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2011.02.010

independent variables (Arciszewski and Ziarko, 1990; Pawlak and Slowinski, 1994; Nguyen and Slezak, 1999; Beynon, 2001; Wan et al., 2009; Slowinski et al., 2009). This paper demonstrates rule extraction by the rough set approach for landslide incidence on the Qingganhe River of the Three Gorges area, China. First, a database of Landsat TM images was assembled by merging several images after geometric corrections. These factors related to landslide occurrence were then edited in a vector spatial database using commercial software (Arc/GIS), including 1:50,000-scale maps of geology, topography, types of slope, and normalized difference vegetation index (NDVI). Following construction of spatial databases in the study area, decision tables were simplified by eliminating superfluous attributes and values of attributes, and by finding simple rules relating environmental factors with landslide incidence. Meanwhile, the relationship between landslide occurrence and environmental factors was statistically analyzed to validate the accuracy of extracted rules by the rough set processing.

2. Study area

The study area covers a region of about 131 km^2 around the Qingganhe River (Fig. 1). The boundary is mainly defined by rivers and ridges, according to the watershed systems and slope properties. As a first-order stream of the Yangtze River, the Qingganhe River flows through the north part of the study area. The north boundary of the area is formed by a mountain ridge adjacent to the Qingganhe River, while the SW and SE boundaries are formed by the Qingganhe River and the Luogudonghe River, respectively.

Geologically, the study area is located on the north-west flank of the Zigui syncline. The syncline has a NNE–SSW trending axis. Geological surveys have indicated that the stratigraphic sequence was not drastically changed by tectonic activities, and hence these strata have generally preserved their original sedimentary sequence. The strata belong to the Triassic to Jurassic basin-fill sediments, which are mostly composed of marly limestone, mudstone, shale, sandstone, and siltstone. These units are very susceptible to landslides and other mass movements.

The study area was partitioned into several cataclinal and anaclinal slopes by the rivers noted above. A detailed landslide inventory map was prepared by means of aerial photographs and extensive field investigations. A total of 27 landslides were identified, as shown in Fig. 2, most of which are located on the cataclinal slopes along the Qingganhe River and the Luogudonghe River. It was noted that the attitude of the strata dip, slope angle, and aspect made a significant contribution to slope instability in the study area (Zeng et al., 2006). The dips of the strata tend to shift from SEE in the northern part to NEE in the southeastern part. The geological balance was disturbed following reservoir construction and the rapid rise in water level, which may have reactivated or initiated new landslides and rock falls. The Qianjiangping landslide, for example, occurred in 2003 after the first stage of reservoir impoundment. The landslide resulted from layered rock with a low dip angle, with a steeply dipping fissure

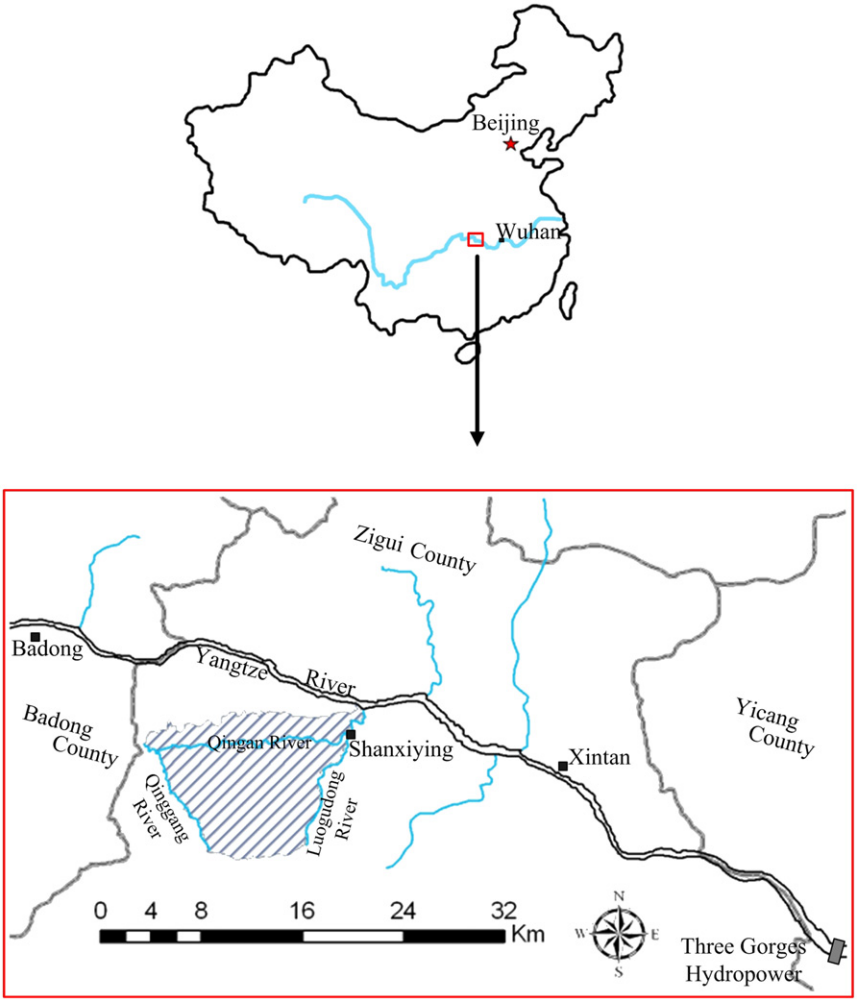

Fig. 1. Geographical location of the study area.