Available online at www.sciencedirect.com
ScienceDirect
Computers, Environment and Urban Systems 32 (2008) 53–65
Computers, Environment and Urban Systems
www.elsevier.com/locate/compenvurbsys

Discerning landslide susceptibility using rough sets

Pece V. Gorsevski [a,*], Piotr Jankowski [b]

[a] *School of Earth, Environment & Society, Bowling Green State University, Bowling Green, OH 43403, USA*
[b] *Department of Geography, San Diego State University, San Diego, CA 92182, USA*

Received 11 September 2006; received in revised form 27 April 2007; accepted 27 April 2007

Abstract

Rough set theory has been primarily known as a mathematical approach for analysis of a vague description of objects. This paper explores the use of rough set theory to manage the complexity of geographic characteristics of landslide susceptibility and extract rules describing the relationships between landslide conditioning factors and landslide events. The proposed modeling approach is illustrated using a case study of the Clearwater National Forest in central Idaho, which experienced significant and widespread landslide events in recent years. In this approach the landslide susceptibility is derived from decision rules of variable strengths computed in rough set analysis and presented on maps for roaded and roadless areas. The rough set approach to modeling landslide susceptibility offers advantages over other modeling methods in accounting for data vagueness and uncertainty and in potentially reducing data collection needs. From an application perspective the rough set-based approach is promising as a decision support tool in forest planning involving the maintenance, obliteration or development of new forest roads in steep mountainous terrain.

Keywords: Landslides; Rough sets; Land-use management; Forest roads; Rule-based predictive models

1. Introduction

Landslides initiated in steep mountainous terrain are a major concern to land-use managers worldwide. Human activities, such as road-building and deforestation accelerate the process of landslides resulting in adverse impacts to the environment (Burton & Bathurst, 1998; Chung, Fabbri, & van Westen, 1995). In the US alone an estimated annual average cost of $1.5 billion dollars due to landslides has been reported (Glade, 1998). During above average wet seasons such as the winter of 2005/06 along the most of U.S. Pacific Coast this number might have been substantially higher. In many developing countries landslides are a serious hazard resulting in losses of life and at least 0.5% of the gross national product (Chung et al., 1995).

Poorly designed land use practices such as road construction and forest harvesting are widely recognized to increase the risk of landsliding in forested and mountainous terrain (Dyrness, 1967; McClelland et al., 1997; Sidle, Pearce, & O'Loughlin, 1985). For instance, roads are often constructed on steep terrain, weak geologic material, and when combined with heavy rainfall, constitute a high-risk situation. Furthermore, when roads are placed on steep slopes, the geometry of the slope is changed because cut slopes are steeper than natural hill slopes. Therefore, roads intercept water flowing downhill altering the natural drainage flow of both surface and subsurface water (Elliot, Foltz, Luce, & Koler, 1996). Change in the forest cover, especially from clear-cuts, unfortunately results in similar consequences as the changes caused by roads, including the tendency to decrease slope stability and increase the risk of landsliding. Landslide events related to forest roads and harvesting are considered as a major reason for deteriorated water quality, loss of fish spawning habitat, and debris jams that may break during peak flows, thereby scouring channels and destroying riparian vegetation.

Therefore, reliable methods of mapping areas susceptible to landslides are essential for land-use management

* Corresponding author.
E-mail addresses: peterg@bgnet.bgsu.edu (P.V. Gorsevski), piotr@geography.sdsu.edu (P. Jankowski).

0198-9715/$ - see front matter
doi:10.1016/j.compenvurbsys.2007.04.001

and are rapidly becoming the standard tool for sound land-use planning. Consequently, there is a need for methods guiding managers to choose the best management strategies while minimizing impacts from land-use activities, such as road construction and forest harvesting in vulnerable slope areas. Many methods and techniques have been proposed to evaluate where or when landslides are most likely to occur, some using Geographic Information Systems (GIS) (Carrara, 1983; Carrara, Cardinali, Guzzetti, & Reichenbach, 1995; Duan & Grant, 2000; Gorsevski, 2002; Gorsevski, Gessler, & Jankowski, 2003; Gorsevski, Gessler, & Jankowski, 2004; Gorsevski, Jankowski, & Gessler, 2005; Gorsevski, Gessler, Foltz, & Elliot, 2006a; Gorsevski, Gessler, Boll, Elliot, & Foltz, 2006b; Gorsevski, Jankowski, & Gessler, 2006c; Hammond, Hall, Miller, & Swetik, 1992; Montgomery & Dietrich, 1994; Okimura & Ichikawa, 1985; Wu & Sidle, 1995). Statistical models that link environmental attributes using spatial correlation are the most widely used methods for mapping landslide susceptibility (Carrara, 1983; Carrara et al., 1991; Chung et al., 1995; Chung & Fabbri, 1999; Dhakal, Amada, & Aniya, 2000; Gorsevski, 2002; Gorsevski et al., 2003, 2004, 2005, 2006a). However, the applicability of statistical methods sometimes is limited by rigid *a priori* data assumptions and the lack of techniques to analyze and characterize the structural relationships existing in the data.

This paper presents an alternative approach to the analysis of landslide susceptibility using the rough set (RS) theory (Pawlak, 1982). The RS theory has been used in a number of discipline-specific applications such as remote sensing (Pal & Mitra, 2002), geographic information science (Ahlqvist, Keukelaar, & Oukbir, 2003), economics (McKee, 2003; Slowinski & Zopounidis, 1995), multi-attribute decision analysis (Pawlak & Slowinski, 1993, 1994), medicine (Komorowski & Øhrn, 1999), civil engineering (Arciszewski & Ziarko, 1990), and artificial intelligence (Predki, Slowinski, Stefanowski, Susmaga, & Wilk, 1998; Predki & Wilk, 1999). The RS theory deals with identifying structural relationships in the data and it is useful in discovering potentially significant facts or data patterns in multidimensional attribute collections. Because of uncertainty and imprecision in classifying information, the RS theory considers information about classification decisions to be vague and approximates information classes by providing their "rough" description (Pawlak & Slowinski, 1993). The approximation in the RS theory reflects different levels of granularity in information while the amount of information affording unambiguous classification of objects determines the degree of roughness. For example, observing landslides from aerial photos with different spatial or spectral resolution will yield different amounts of information for classification of landscape features into these associated with the landslides and those not associated with landslides.

Compared to other mathematical approaches that deal with vagueness and uncertainty the RS theory bears some resemblance to the Dempster–Shafer (D–S) theory (Dempster, 1967; Shafer, 1976). However, the difference between the two theories is that the RS theory uses sets of lower and upper approximations to represent knowledge in data collections while the D–S theory uses belief functions represented by lower and upper probability functions. The approximations for a given data set derived with the RS theory are based solely on data while the approximations derived by the D–S theory involve calculations of belief values using both subjective judgments and data (Dempster, 1967).

The D–S methodology, coupled with the fuzzy k-means, was used by Gorsevski et al. (2005) to predict road related (RR) landslide susceptibility (spatial locations within roaded areas) and non-road related (NRR) susceptibility (spatial locations within non-roaded areas) for a study site used by McClelland et al. (1997). This paper focuses on the application of the RS methodology to the same study area using the same datasets and comparatively evaluates the fitness of both methodologies to map landslide susceptibility. The proposed approach is demonstrated using a case study of the Clearwater National Forest (CNF) in central Idaho. The study purpose has been to address the following two research questions:

(1) Can RS models for RR and NRR landslide susceptibility be developed with the same predictor variables?
(2) Is spatial prediction of landslide susceptibility using the RS approach better than spatial prediction using the D–S approach?

Answering the first question will help establish whether different combinations of predictor variables are necessary in the development of predictive RS models and if the development of two independent models for RR and NRR landslide susceptibility is necessary. Answering the second question will help determine if the RS methodology yields better predication of landslide susceptibility than the D–S approach.

2. Rough set theory

The rough set (RS) theory was introduced by Pawlak (1982, 1991) as a mathematical framework for approximate reasoning dealing with uncertainty and vagueness in decision making processes. The theory belongs to a branch of computer science called soft computing and has been used in data mining, knowledge discovery, pattern recognition, machine learning, and other areas of artificial intelligence. The RS theory is based on the assumption that each object in the universe is associated with knowledge which can be used to classify it. The knowledge is represented in an *information table or an information system* (data table) where rows represent objects (for example, landslide locations represented by polygons or points) and the columns represent attributes (for example, elevation, slope, solar radiation, and wetness index). A special form of information table is called a *decision table*, where one column represents