Geomorphology 201 (2013) 45–51

Contents lists available at ScienceDirect

Geomorphology

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

A comparison between Bayes discriminant analysis and logistic regression for prediction of debris flow in southwest Sichuan, China

Wenbo Xu [a,*], Shaocai Jing [a], Wenjuan Yu [a], Zhaoxian Wang [a], Guoping Zhang [b], Jianxi Huang [c]

[a] School of Resources and Environment, University of Electronic Science and Technology of China, 611731 Chengdu, Sichuan Province, China
[b] Public Weather Service Center of China Meteorological Administration, 100081 Beijing, China
[c] College of Information and Electrical Engineering, China Agricultural University, 100083 Beijing, China

ARTICLE INFO

Article history:
Received 19 January 2013
Received in revised form 4 June 2013
Accepted 7 June 2013
Available online 15 June 2013

Keywords:
Debris flow
Rainfall and environmental factors
Prior probability
Bayes discriminant analysis
Logistic regression

ABSTRACT

In this study, the high risk areas of Sichuan Province with debris flow, Panzhihua and Liangshan Yi Autonomous Prefecture, were taken as the studied areas. By using rainfall and environmental factors as the predictors and based on the different prior probability combinations of debris flows, the prediction of debris flows was compared in the areas with statistical methods: logistic regression (LR) and Bayes discriminant analysis (BDA). The results through the comprehensive analysis show that (a) with the mid-range scale prior probability, the overall predicting accuracy of BDA is higher than those of LR; (b) with equal and extreme prior probabilities, the overall predicting accuracy of LR is higher than those of BDA; (c) the regional predicting models of debris flows with rainfall factors only have worse performance than those introduced environmental factors, and the predicting accuracies of occurrence and nonoccurrence of debris flows have been changed in the opposite direction as the supplemented information.

1. Introduction

Debris flow is a sudden natural disaster specifically occurring in mountain areas, with strong carrying, lashing, and burying abilities and intense destructiveness, and has become a huge threat to the security of human life and property and an obstacle to economic development (Ma, 2010). The occurrences of debris flows owe to the interaction of geology, topography, geomorphology, hydrology, weather, and other natural factors, which can be divided into two groups: one is rainfall, which directly triggers the occurrence of debris flow, and the other is the environmental factors that are the basic conditions of the occurrence of debris flow. This disaster has caught unprecedented attention in the world; lots of researchers are continuously carrying out relevant research, mainly focusing on its prediction.

In the earlier debris flow prediction, most prediction models were built by means of the investigation of relationship between rainfall and debris flow on the basis of the processing of rainfall data (Kenneth, 1987; Tan and Han, 1992; Chen et al., 2007; Shieh et al., 2009). With the deep-going research of debris flow prediction and the innovative development of data obtaining technologies, environmental factors are paid more attention. These environmental factors are comprehensively analyzed in order to conduct debris flow susceptibility evaluation and risk zoning (Lee, 2005; Liu, 2006; Pradhan and Lee, 2007; Pradhan, 2010). And these factors—along with rainfall factors—are used as independent predictor variables for debris flow forecasting; that is, the forecasting model is established containing these two kinds of parameters (Jomelli et al., 2003; Ohlmacher and Davis, 2003; Rupert et al., 2008). Therefore, considering the environmental factors for the prediction of debris flow is necessary.

Recently, the prediction of debris flow mainly focuses on two aspects: mechanical prediction based on the disaster formation mechanism (Liu, 2002) and quantitative prediction based on mathematical statistics. The quantitative prediction models are usually applied in the research of regional debris flow, and then this kind of prediction can be divided into probabilistic prediction and deterministic prediction according to the predicting results. Probabilistic prediction is represented by the logistic regression (LR) model, which has been used to build a model based on the combination of various rainfall and environmental factors (Ohlmacher and Davis, 2003; Rupert et al., 2008). Deterministic prediction is represented by the Bayes discriminant analysis (BDA); Spiegelhalter (1986) applied the Bayes formula to build spatial forecasting models at an earlier time, and later this method was used for the prediction of debris flows and landslides (Leclerc, 1994; Graf et al., 2009).

The multivariate statistical methods, BDA and LR, are widely used for analysis of data in event classification. Many researchers have used two classification methods in various practical fields (Maja et al., 2004; Alkarkhi and Easa, 2008), especially in health sciences and clinical psychology (Payne et al., 1998; Udris et al., 2001). The LR is a form of regression and uses the logit transformation to calculate the ratio of probability by using probability outcome divided by probability without it and to predict the probabilities of group memberships in relation to several variables (Worth and Cronin, 2003). Bayes discriminant analysis is derived from the linear discriminant analysis (LDA),

* Corresponding author. Tel./fax: +86 28 61831571.
E-mail address: xuwenbo@uestc.edu.cn (W. Xu).

0169-555X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.geomorph.2013.06.003

which distinguishes new samples and classifies them into known groups (Fan and Mei, 2002). The two methods have different basic ideas; in the whole, BDA is usually used to simulate the linear relationships between the independent variables and dependent variables under the assumptions of multivariate normality and equal covariance, while LR simulates the nonlinear relationship and makes no such assumptions (Lei and Koehly, 2003). In general, BDA will give better results when these assumptions are met, but in other cases LR will be more suitable (Efron, 1975; Harrell and Lee, 1985). However, which of these two methods will be selected is more relevant to the actual statistical application field than to the assumption satisfaction. In practice, the assumptions are nearly violated, therefore doing continuous simulation experiments to find the more suitable method is necessary (Maja et al., 2004). In addition, prior probabilities, which are the proportions of group members that exist in the populations, also affect the classification results of BDA and LR. For instance, Fan and Wang (1999) and Lei and Koehly (2003) compared the classification error rates of LDA and LR by using the Monte Carlo simulation under different prior probabilities in the binary cases. Consequently, both methods are very applicable in debris flow prediction and worthy of study.

In view of the previous research, the feasibility of modeling based on the occurrence mechanism is low; and the mathematical statistical model still occupies an important position in the prediction of regional debris flow. Nevertheless, BDA is less used than LR for debris flow prediction, and the studies of comparison between the two are much rarer. On the mastery of both theoretical methods, the main objective of this study is to compare the performances of BDA and LR to predict debris flow with different combinations of debris flow prior probabilities in terms of the historical debris flow data in the period 1981–2000, including rainfall and environmental background data.

2. Study areas and data source

2.1. Study areas

According to the susceptibility of debris flows in Sichuan Province (Xu et al., 2013), Panzhihua and Liangshan Yi Autonomous Prefecture belonging to the debris-flow-prone areas, are located in the southwest of Sichuan and are bounded by longitudes of 100°15′ E. and 103°53′ E., latitudes of 26°03′ N. and 29°27′ N (Fig. 1). Obvious dry and wet seasons are to provide concentrated rainfall. Elevation has significant differences, a low-lying West High East; and the complex topography is mainly mountainous. The geological structure is also complicated, with staggered fault zones and seismic belts. Hence, debris flow is easy to outbreak in the regions.

2.2. Data sources

Firstly, the corresponding debris flow material, in the period 1981–2000, were extracted from the China Institute of Geo-Environmental Monitoring, which contain attributes about event time and accurate positioning information like latitude and longitude. And these data were based on a day as the event unit. The count was 129. In order to meet the model condition (which is that building the models needs occurrence and nonoccurrence of debris flow), the precipitation records of meteorological stations that were from the nearest disaster points

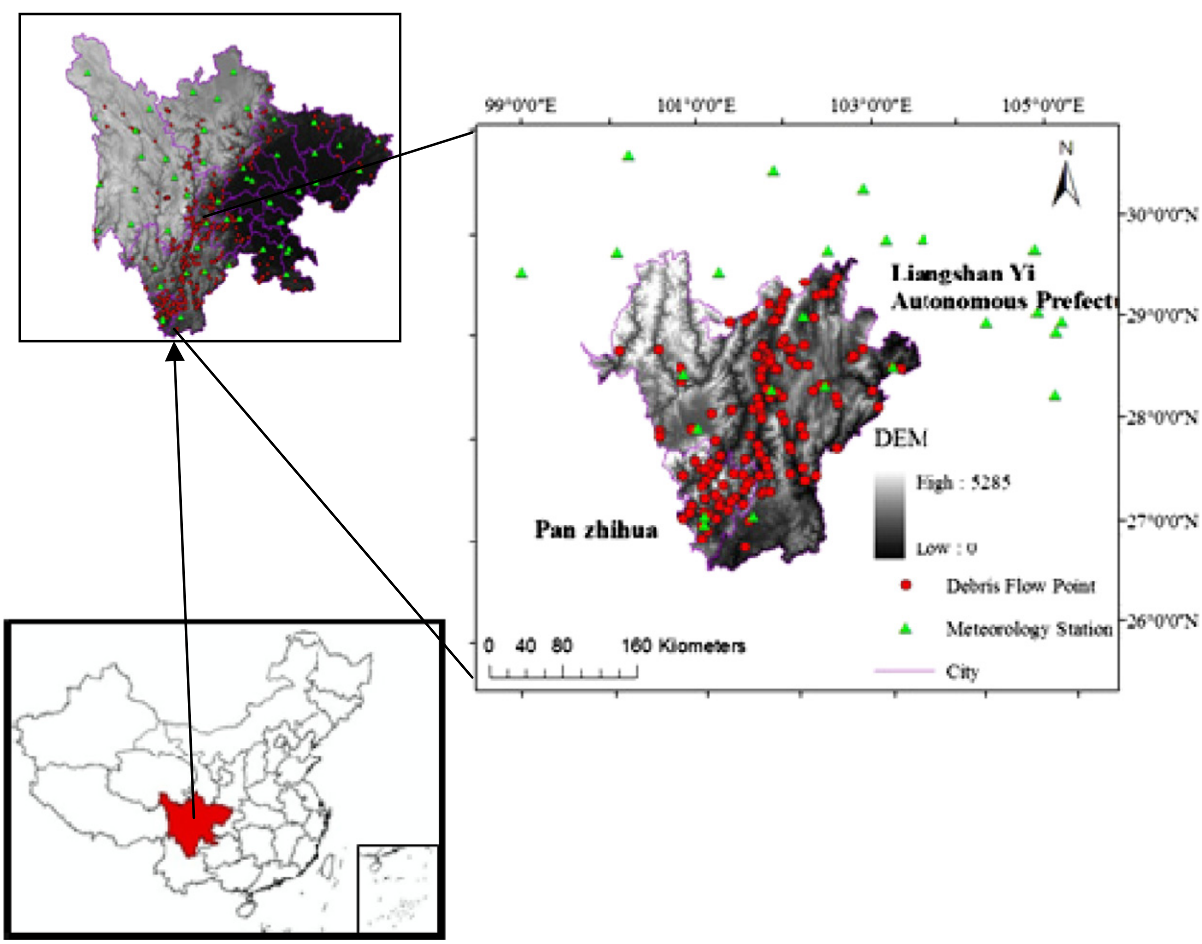

Fig. 1. Map of distribution of Liangshan Yi Autonomous Prefecture and Panzhihua Administrative Region, debris flow (1981–2000) and DEM.