Catena 166 (2018) 317–327

Contents lists available at ScienceDirect

Catena

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

A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping

A-Xing Zhu[a,b,c,d], Yamin Miao[a,b,c], Rongxun Wang[e], Tongxin Zhu[e], Yongcui Deng[a], Junzhi Liu[a,b,c,*], Lin Yang[f], Cheng-Zhi Qin[f], Haoyuan Hong[a,b,c,**]

[a] Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Nanjing 210023, China
[b] State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China
[c] Jiangsu Center for Collaborative Innovation in Geographic Information Resource Development and Application, Nanjing, Jiangsu 210023, China
[d] Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA
[e] Department of Geography, University of Minnesota-Duluth, Duluth, MN, USA
[f] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Keywords:
Data-driven models
Expert knowledge-based model
Landslide susceptibility
GIS
Logistic regression
Artificial neural network

ABSTRACT

In this study, an expert knowledge-based model, a logistic regression model, and an artificial neural network model were compared for their accuracy and portability in landslide susceptibility mapping. Two study areas (the Kaixian and the Three Gorges areas in China) were selected for this comparison based on their well-known, high landslide hazard. To evaluate the performance of these models and to minimize the impact of the particularity of a study area on model performance, cross-applications of three models between the two study areas were conducted. When the Kaixian area was used as a model development area, prediction accuracy for the expert knowledge-based model, the logistic regression model, and the artificial neural network model were 71.5%, 81.0% and 100.0%, respectively. The high prediction accuracy of the two data-driven models were expected because the observed landslide occurrence used in training the models were also used to validate their respective performance, while the expert knowledge-based model did not use these observations for training. The perfect accuracy for the neural network model can also be attributed to its over-prediction of the susceptibility. When breaking the susceptibility into four classes: very low susceptibility (0–0.25), low susceptibility (0.25–0.5), high susceptibility (0.5–0.75), and very high susceptibility (0.75–1), the observed landslide density at the very high susceptibility level is 0.303/km^2, 0.212/km^2, and 0.195/km^2 for the expert knowledge-based model, the logistic regression model, and the artificial neural network model, respectively. This suggests that the expert knowledge-based model was much better than the other two data-driven models at evaluating landslide occurrence in very high susceptibility areas. When the three models developed in the Kaixian area were applied in the Three Gorges area without any changes, their prediction accuracy dropped to 44.8% for the logistic regression model and 81.6% for the artificial neural network model, while the expert knowledge-based model maintained its initial accuracy level of 82.8%. The landslide density for the very high susceptibility areas in the Three Georges area was 0.275/km^2, 0.082/km^2, and 0.060/km^2 for the expert knowledge-based model, the logistic model, and the artificial neural network model, respectively. These results indicate that the expert knowledge-based model is more effective at predicting areas with very high susceptibility. When the Three Gorges area was used as a model development area and the Kaixian area was used as the model application area, similar results were obtained. Results from the two experiments show that the performance of the logistic regression model and artificial neural network model is not as stable as the expert knowledge-based model when transferred to a new area. This suggests that the expert knowledge-based model is more suitable for landslide susceptibility mapping over large areas.

* Correspondence to: J. Liu, Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Nanjing 210023, China.
** Correspondence to: H. Hong, State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China.
E-mail addresses: liujunzhi@njnu.edu.cn (J. Liu), hong_haoyuan@outlook.com (H. Hong).

https://doi.org/10.1016/j.catena.2018.04.003
Received 22 December 2017; Received in revised form 22 February 2018; Accepted 3 April 2018

1. Introduction

Landslides are the most common geological hazards and result in substantial human casualties and property losses (Lee and Choi, 2004). In order to mitigate damages caused by landslides, many models have been developed to map their potential distribution (Atkinson and Massari, 1998; Brenning, 2005; Xu et al., 2012a, 2012b; Ozdemir and Altural, 2012). In spite of different implementation procedures, almost all models attempt to evaluate landslide occurrence based on the general principle that future landslides will be more likely to occur under environmental conditions that are the same as they were for past landslide events (Varnes, 1984; Carrara et al., 1995; Guzzetti et al., 1999). Landslide susceptibility is depicted as a function of predisposing factors, and a susceptibility value can be predicted as long as the predisposing factors and the relationship between landslide susceptibility and predisposing factors are known.

Data-driven models, including statistical methods and machine learning algorithms, have been commonly applied in landslide susceptibility mapping and have demonstrated high predictive capability and robustness, assisted by the development of spatial information technologies (Lee et al., 2003; Ermini et al., 2005; Bai et al., 2010, 2012; Pradhan et al., 2010). The data-driven models acquire and analyze the relationship between landslide susceptibility and predisposing factors objectively from past landslide occurrence and predisposing factors at those sites to predict areas prone to landslides (Süzen and Doyuran, 2004; Wang and Sassa, 2005; Ayalew and Yamagishi, 2005). Among the many data-driven models, logistic regression model and artificial neural network model are among the most widely used.

Zhu et al. (2014) have proposed an expert knowledge-based model to map landslide susceptibility. The model extracts knowledge of the complicated nonlinear relationship between landslide susceptibility and predisposing factors from local domain experts by using knowledge acquisition techniques. The relationship is represented as a set of fuzzy membership functions rather than relationship in the form of a logistic regression or weights in an artificial neural network mined from landslide occurrence (Zhu et al., 2004, 2014). This expert knowledge-based model has been shown to be effective in predicting landslide susceptibility and to be portable when applied to a new area with landslide mechanisms that are similar to those of the model development area without changing the knowledge base (Zhu et al., 2014).

Even though these models have been successful and applied widely, one question that needs to be examined is how they compare with each other in terms of both their ability to evaluate susceptibility and their portability to areas outside those in which they were developed. One way to evaluate model performance is to carry out cross-applications (Pradhan et al., 2010). Few such studies of these models have been compared. Furthermore, the expert knowledge-based model has never been compared with other models to evaluate its performance.

The aim of this study is to compare the expert knowledge-based model with the data-driven models with respect to model performance and portability. In order to achieve these goals, the logistic regression model and artificial neural network model, which have been widely used and demonstrated to provide reliable results of landslide susceptibility, were chosen to represent data-driven models. Two study areas (the Kaixian study area and the Three Gorges study area in China) were selected because of their high landslide hazard and their similar geomorphological settings and landslide mechanisms. Cross-applications of the three models between the two study areas were conducted to test their respective performance and portability. The comparative study of these models not only provides reference for disaster mitigation and regional planning, but also adds to the literature on landslide susceptibility mapping.

2. Study areas and materials

Two areas located in the middle-upper reach of the Yangtze River in China, the Kaixian and Three Gorges study areas (Fig. 1), were used to perform the expert knowledge-based model, the logistic regression model, and the artificial neural network model. The areas are recognized as having a high, natural risk for landslides (Wu et al., 2001; Liu et al., 2004; Zhu et al., 2004, 2014).

2.1. Study areas

The Kaixian study area, with a total area of about 250 km^2, is located in Kaixian County (Fig. 1), Chongqing Municipality. Its elevation ranges from about 140 m to about 1070 m above sea level, with an average elevation of 390 m. The greatest local relief is about 700 m and the average is about 300 m. Most of the slopes are very steep with an average gradient of about 20°. The lithology of this area consists of three major types: the lower to middle Jurassic system made of sandstone, siltstone, mudstone, and shale; the upper Jurassic system composed of sandstone and siltstone; and the Quaternary system composed mostly of recent deposits along the river valleys (Zhu et al., 2014).

The Three Gorges study area is situated along the Yangtze River in the Chongqing Municipality between Yunyang county and Wushan county, with a total area of about 4600 km^2 (Fig. 1). The lithology of the area comprises two major types: the Jurassic system composed of mudstone, sandstone, siltstone, shale, and coal; and the Triassic system composed of limestone, shale, claystone, dolomite, gypsum, sandstone, siltstone, and coal. The geomorphological settings and landslide mechanisms of the Three Gorges study area are similar to these of the Kaixian study area except for some different lithology types (Zhu et al., 2014).

2.2. Landslide inventory data

The landslides included in the study needed to have occurred after the field data were collected for the predisposing environmental factors, since the occurrence of landslides often dramatically changes a landform (Conoscenti et al., 2015). The altered landform cannot reflect the conditions prior to landslide occurrence. A topographic map created in 1960s and a geology map created in 1970s were used. To ensure the eligibility of landslide events, only those landslides that took place after the 1970s were selected for the study. In addition, landslide sliding surface is in reality an area, and not a single point, but, it is very difficult to determine the exact extent of a landslide. To avoid the uncertainty introduced by the boundary issue, every landslide occurrence was represented by one single point (or one single pixel in the raster representation). The location of this point was then given by local landslide experts based on their field observation. This approach can reduce error related to the exact location of the borderline and has been used in several studies (Atkinson and Massari, 1998).

Based on the principles proposed above, 21 landslide occurrences since 1978 were recorded for the Kaixian study area and 205 for the Three Gorges study area, occurrences that were determined with the help of local landslide experts. The landslide occurrence in both study areas were used to develop the logistic regression and the artificial neural network models and cross-applied to validate the respective performance of the three models. The expert knowledge-based model was developed based on the expertise of local experts. The field-observed landslide occurrences were not used in developing the expert knowledge-based model.