Landslide Susceptibility Analysis and its Verification using Likelihood Ratio, Logistic Regression and Artificial Neural Network Methods: Case study of Yongin, Korea

S. Lee
Geoscience Information Center, Korea Institute of Geology & Mineral Resources (KIGAM)
30, Gajung-dong, Yusung-gu, Daejeon, 305-350, Korea
leesaro@kigam.re.kr

J. H. Ryu
Ocean Climate and Environmental Research Division, Korea Ocean Research and Development Institute (KORDI)
Ansan P.O. Box 29, Seoul 425-600, Korea
jhryu@kordi.re.kr

Abstract: The likelihood ratio, logistic regression and artificial neural networks methods are applied and verified for analysis of landslide susceptibility in Yongin, Korea using GIS. From a spatial database containing such data as landslide location, topography, soil, forest, geology and land use, the 14 landslide-related factors were calculated or extracted. Using these factors, landslide susceptibility indexes were calculated by likelihood ratio, logistic regression and artificial neural network methods. Before the calculation, the study area was divided into two sides (west and east) of equal area, for verification of the methods. Thus, the west side was used to assess the landslide susceptibility, and the east side was used to verify the derived susceptibility. The results of the landslide susceptibility analysis were verified using success and prediction rates. The verification results showed satisfactory agreement between the susceptibility map and the existing data on landslide locations.
Keywords: Landslide Susceptibility, Likelihood ratio, Logistic Regression, Artificial Neural Network, Korea

1. Introduction

In Korea, landslides are a recurrent problem throughout most of the country. They cause extensive damage to property and occasionally result in loss of life. In particular, personal injury and damage to property in 1991, 1996, 1998, 1999 and 2002 was great. Most of the landslides are triggered by heavy rainfall in Korea, but few attempts are made to predict them or to prevent damage. To remedy this, it is necessary to scientifically predict which areas are susceptible to landslides. For the landslide-susceptibility analysis of the Korean situation, likelihood ratio, logistic regression and artificial neural networks methods were applied and verified for the study area of Yongin, Korea. As the basic analysis tool, a Geographic Information System (GIS) was used for spatial data management and manipulation.

For application and verification of landslide susceptibility methods, the study area was divided into two sides, west and east, of equal area. The west side was used to assess the landslide susceptibility using the methods of this work and the east side was used to verify the methods that were applied to the west side. Landslide occurrence areas were detected in the study area by interpretation of aerial photographs and field surveys. Topography, soil, forest, geology, and land use spatial databases were constructed for the analysis. Using the detected landslide locations and the calculated or extracted factors, three landslide analysis methods were applied: likelihood ratio, logistic regression, and artificial neural networks. For the application of these, a statistical and an artificial neural network program were used with a GIS program. Finally, the analysis results were verified using data from not only the east side, but also the west side of the study area..

2. Study area and spatial database

The Yongin study area had high landslide damage after heavy rain in 1991 and was selected as a suitable case to evaluate the frequency and distribution of landslides. The site lies between the latitudes 37.14° N and 37.19° N, and longitudes 127.11° E and 127.23° E, and covers an area of 66 km^2. In the study area, the landslides were mainly debris flows and shallow soil slips that occurred during 3–4 hours of high intensity rainfall, or shortly afterwards. The landslides occurred where the maximum daily rainfall exceeded 114 mm (Lee and Min 2001).

Identification and mapping of a suitable set of instability factors (thematic mapping) bearing a relationship with slope failures requires an *a priori* knowledge of the main causes of landslides (Guzzetti *et al.,* 1999). These instability factors include surface and bedrock lithology and structure, bedding altitude, seismicity, slope steepness and morphology, stream evolution, groundwater conditions, climate, vegetation cover, land use, and human activity. The availability of thematic data varies widely, depending on the type, scale, and method of data acquisition.

In order to apply the landslide susceptibility analysis method in the study area, a spatial database of landslide-related maps and images, such as topography, soil, forest, geology and land use was designed and constructed, and the landslide-related factors were calculated or extracted from the maps and images. A map of recent landslides was developed from 1:20,000 scale aerial photographs, in combination with the GIS, and this was used to evaluate the frequency and distribution of landslides in the area. Maps and imagery relevant to landslide occurrence were constructed and processed for the spatial database using the GIS software ARC/INFO. These included 1:5,000 scale topographic maps, 1:25,000 scale soil maps, 1:25,000 scale forest maps, 1:50,000 scale geological maps, and Landsat TM satellite imagery.

There are 14 factors considered in calculating the susceptibility index by likelihood ratio, logistic regression, and artificial neural network analysis. The factors were calculated or extracted from the constructed spatial database. In the process, because all the factors were converted to ASCII data for application of statistical and artificial neural networks, a vector-to-raster conversion was undertaken to provide a raster layer of landslide areas, with 10 m × 10 m cells. Contour and survey base points, which have an elevation value read from the topographic map, were extracted, and a Digital Elevation Model (DEM) was made with 10 m resolution. Using the DEM, the slope, aspect and curvature were calculated. Texture, topography, drainage, material, and thickness of soil were acquired from a soil map, and forest type, stand diameter, age and density were obtained from forest maps. The lithology map was obtained from a geological map and land use data were obtained from LANDSAT TM imagery.

3. Landslide susceptibility analysis

1) Likelihood ratio method

The likelihood ratio is calculated from analysis of the relation between landslides and the relevant factors. Therefore, the likelihood ratios of each factor's type or range were calculated from their relationship with landslide events to the west side of the study area. In the relation analysis, the ratio is that of the area where landslides occurred to the total area, so that a value of 1 is an average value. The 14 landslide-related factors are slope, aspect, curvature and type of topography, texture, drainage, material, thickness of soil, forest type, stand age, stand diameter, stand density, lithology and land use. The factors were converted to a raster grid with 10 m × 10 m cells for application of the likelihood ratio method. To calculate the Landslide Susceptibility Index (LSI), each factor's likelihood ratio values were summed to the west side of the study area as in Eq. (1).

$$LSI = \Sigma Fr \quad (1)$$

(LSI: Landslide Susceptibility Index; Fr: Rating of each factors' type or range)

2) Logistic regression method

Logistic regression, which is one of the multivariate analysis methods, is useful for predicting the presence or absence of a characteristic or outcome based on values of a set of predictor variables. Using the logistic regression method, the spatial relationship between landslide-occurrence location and landslide-related factors was calculated. In addition, logistic regression formulas were created as shown in formulas (2) and (3). Finally, the probability which predicts the possibility of landslide-occurrence, for the west side of the study area, was calculated using spatial database, the coefficients and Eqs. (2) and (3):

$$z = (0.035 \times SLOPE) + (-0.030 \times CURVA) + ASPECT_b + TOPO_b + TEXTURE_b + DRAIN_b + SMATER_b + THICK_b + SANG_b + KUNG_b + YUNG_b + FMILDO_b + GEOL_b + LANDUSE_b - 15.931 \quad (2)$$

$$p = \exp(z) / (1 - \exp(z)) \quad (3)$$

(where Slope is slope value; Curva is curvature value; $ASPECT_b$, $TOPO_b$, $TEXTURE_b$, $DRAIN_b$, $SMATER_b$, $THICK_b$, $SANG_b$, $KUNG_b$, $YUNG_b$, $FMILDO_b$, $GEOL_b$, $LANDUSE_b$ are logistic regression coefficient values; z is a parameter; and p is the probability of landslide occurrence)

3) Artificial neural network method

The purpose of an artificial neural network is to build a model of the data-generating process so that the network can generalize and predict outputs from inputs that it has not previously seen. Artificial neural networks are trained by the use of a learning rule and a set of examples of associated input and output values. Landslide susceptibility was analyzed using an artificial neural network program that was partially modified and upgraded from the original version developed by Hines (1997) in the MATLAB package.

4. Verification of landslide susceptibility analysis methods

1) Method

The landslide susceptibility analysis was performed using likelihood ratio, logistic regression, and artificial neural network methods for the west side of the study area, and the analysis results verified using the landslide locations for the east side study area. The verification method was performed by comparison of existing landslide data and landslide susceptibility analysis results for the west side of the study area. The success rates illustrate how well the estimators perform with respect to the west side landslides used in constructing those estimators. The prediction rates on the other hand, are used as measurements of how well the probability model and its estimators predict the distribution of future landslides (Chung and Fabbri, 1999).

2) Success rate

The success rate verification results are divided into 20 classes of accumulated area ratio % according to the landslide susceptibility index value. For example, the 75 to 100% class contains 58.4% of the west side of the study area in success rate using the likelihood ratio method. In addition, it occupies 58.0% of the west side of the study area in success rate using the logistic regression method, and 48.3% of the west side of the study area in success rate using the artificial neural network method. So, the value of 58.4% in the likelihood ratio method is better than the 58.0% in the logistic regression, but the 48.3% in the artificial neural network method is worse than 50.8% in the logistic regression. Hence, the likelihood ratio method fits best in the 75–100 class. Although, for the first four classes (70 to 100%), the likelihood ratio method is little better than those from the logistic regression method except for one class (90 to 95%). For the remainder of the classes (0–70%), the logistic regression method produced somewhat better results than the likelihood ratio method. The artificial neural network method was worse than the likelihood ratio and logistic regression methods in all classes.

3) Prediction rate

The prediction rate verification results are divided into 20 classes with accumulated area % according to landslide susceptibility index value. For example, the 75 to 100% class contains 49.1% of the east side of the study area in prediction rate for the likelihood ratio method. In addition, it occupies 42.4% of the west side of the study area in prediction rate for the logistic regression method and 51.8% of the east side of the study area in prediction rate for the artificial neural network method. The value of 49.1% for the likelihood ratio is better than the 42.4% for logistic regression, but the 51.8% for the artificial neural network method is better than 49.1% for the likelihood ratio. Hence, the artificial neural network method fits best in the 75 to 100% class.

The success rates for the first two classes (90 to 100%) are better than those for the artificial neural network method. For the remainder of the classes (65–90%) and the middle classes (40–65%), the artificial neural network method produced better results than the likelihood ratio method. The logistic regression is better than the other two methods below the 35% classes.

5. Conclusion and discussion

The likelihood ratio method is simple and the process of input, calculation and output could be understood easily. Moreover, because the likelihood ratio value can be used as a rating, there is no need to convert the database to another format, such as ASCII. A large amount of data can be processed in a GIS environment quickly and easily. The logistic regression and artificial neural network methods require that the data be converted to ASCII for use in the statistical package and the artificial neural network program and later reconverted to incorporate it into the GIS database. Moreover, the large amount of data cannot be processed in the statistical package and artificial neural network program quickly and easily. In addition, if the factors are discrete data, the integer values are re-assigned according to the discrete data range or type. However, correlation of landslide and other factors can be analyzed qualitatively.

For the method to be more generally applied, more landslide data are needed, as well as being applied to more regions. Fortunately, the landslide-related spatial database including topography, soil, forest, geology and land use that were used in the study, is available for most areas of Korea already, so the landslide analysis can be done systematically and quickly for all of Korea. Landslide susceptibility maps are of great help to planners and engineers for choosing suitable locations to implement developments. These results can be used as basic data to assist slope management and land-use planning, but the methods used in this study are valid for generalized planning and assessment purposes, although they may be less useful at the site-specific scale where local geological and geographic heterogeneities may prevail.

References

[1] Chung C.F and Fabbri A.G.., 1999, Probabilistic prediction models for landslide hazard mapping. *Photogrammetric Engineering & Remote Sensing, 65,* 1389-1399.

[2] Guzzetti F., Carrarra A., Cardinali M. and Reichenbach, P., 1999, Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology, 31,* 181-216.

[3] Hines J. W., 1997, *Fuzzy and Artificial neural Approaches in engineering.* John Wiley and Sons, Inc.

[4] Lee S. and Min K., 2001, Statistical analysis of landslide susceptibility at Yongin, Korea. *Environmental Geology, 40,* 1095-1113.