VALIDATION OF LOGISTIC REGRESSION MODEL FOR LANDSLIDE SUSCEPTIBILITY MAPPING AT GANEOUNG AREAS, KOREA

HYUN-JOO OH AND SARO LEE

Geoscience Information Center, Korea Institute of Geoscience & Mineral Resources (KIGAM), 92 Gwahang-No, Yuseong-Gu, Daejeon, 305-350, Korea

E-Mail: leesaro@kigam.re.kr, Tel:+82-42-868-3057, Fax:+82-42-868-3413

1. INTRODUCTION

Landslides as major natural geological hazards often result in significant harm to people and property. Between 20 and 31 August, 2002, Typhoon Rusa hit the Gangneung area as a storm associated with heavy rainfall. The maximum daily rainfall amounted to 609 mm, with 80 mm per hour. The consequence was the death of 266 people. Damage to property amounted to an approximate value of US$ 8 billion. Amongst the casualties 81 people died as a result of landslide and the collapse of cut slope faces. So, the purpose of the study is to detect the landslide location using satellite images, and landslide-susceptibility analysis techniques were applied and validated using a logistic regression model. Using GIS as the basic analysis tool for landslide hazard mapping can be effective for management and manipulation of spatial data, together with some reasonable models for the analysis. Recently, there have been studies on landslide hazard evaluation using GIS, and many of these studies have applied statistical models available, the logistic regression model, has been applied to landslide hazard mapping. [1][2][3][4] In this study, GIS software, Arcview 3.3 and ArcGIS 9.2, and statistical software, SPSS 12.0 was used as the basic analysis tool for spatial management and data manipulation.

2. DATA AND METHODOLOGY

For probabilistic landslide hazard analysis, accurate detection of the location of landslides is very important. The application of remote sensing methods, such as aerial photographs and satellite imagery, is used to obtain significant and cost-effective information on landslides. So, in this study, 6.6m-resolution KOMSAT EOC (Korea Multipurpose Satellite Electro-Optical Camera) images were used. The KOMSAT-1 ECO images used to detect the landslide are December 7, 2001 (taken before the landslides occurred) and December 14, 2002 (after the landslides occurred). For change detection, we used the image difference method (Red Green Difference Image). We exclude the changes that are not related to landslides, and validated the result by a field survey. Importantly, we knew from the field survey that there was much flooding near the stream that might be interpreted as a landslide. Therefore, we used the 1:5,000 scale digital topographic map to overlay with the images and exclude the stream flooding. Then the selected area was check by the field survey using GPS and 1:5,000 scale topographic map. In total, 457 landslides were mapped within a total study area of 47.94 km^2 in Sagimakri. In total, 116 landslides were mapped within a total study area of 9.70 km^2 in Samgyori. In this study area, the most of landslide is soil slide and debris flow.

Topography, soil, forest, geology and land cover databases were constructed as part of the analysis. Topographic factors such as altitude, slope, aspect, curvature and drainage were calculated from the topographic database. Soil texture, material, drainage, effective thickness and topographic type were extracted from the soil database. Forest type, forest diameter and forest density were extracted from the forest database. Lithology was extracted from the geological database and land cover data was classified from Landsat TM images. Lineaments were extracted from the IRS image and aerial photographs by visual interpretation by a structural geologist, and the proximity of landslides to the lineament was measured. Using the detected landslide locations and the constructed spatial database, a landslide susceptibility analysis method, spatial logistic regression, was applied and validated. For this, the calculated and extracted factors were converted to a 5 m × 5 m grid (ARC/INFO GRID type). The grid data was converted into ASCII file and then imported to the statistical program used. Then, using a logistic regression model, the spatial relationships between the landslide location and each landslide-related factor were analyzed and a formula of landslide occurrence possibility was extracted using the relationships in the statistical program. This formula was used to calculate the landslide susceptibility index and was mapped using the grid. The susceptibility map was validated using existing landslide locations. For cross-validation, the formula was also applied to another study area, Sagimakri in Korea, to calculate and map the landslide susceptibility index there. The same spatial database and factors for the Sagimakri area as for Samgyori were available to us. Finally, the susceptibility map was again validated using the existing landslide location.

3. RESULT

A statistical program was used to calculate the correlation of a landslide event to each factor. Firstly, all factors were constructed in the database and then logistic regression coefficients of the factors were calculated. The coefficients of the logistic regression model were estimated using the maximum-likelihood method. In other words, coefficients that make the observed results most likely are selected. Since the relationship between the independent variables and the probability is nonlinear in the logistic regression model, an iterative algorithm is necessary for parameter estimation. In case of Sagimakri area, there are positive associations, such as slope and distance from drainage and negative associations, such as curvature. In case of Samgyori area, there are positive associations, such as slope and negative associations, such as curvature and distance from drainage. After interpretation, formulas (1) and (2-3), which predict the landslide-occurrence possibility, were created.

$$p = 1 / (1 + e^{-z}) \text{ or } p = e^{z} / (1 + e^{z}) \quad (1)$$

Sagimakri

$$z = (0.109 \times SLOPE) + (0.002 \times DISTDRAIN) + (-0.004 \times DISTLINEA) + ASPECTw + CURVAw + TOPOw + TEXTUREw + DRAINw + MATERIALw + THICKw + TYPEw + DIAMETERw + AGEw + DENSITYw + GEOLw + LANDUSEw - 33.173 \quad (2)$$

Samgyori

$$z = (0.109 \times SLOPE) + (-0.005 \times DISTDRAIN) + (-0.009 \times DISTLINEA) + ASPECTw + CURVAw + TOPOw + TEXTUREw + DRAINw + MATERIALw + THICKw + TYPEw + DIAMETERw + AGEw + DENSITYw + GEOLw + LANDUSEw - 33.173 \quad (3)$$

where SLOPE is slope value; DISTDRAIN is distance from drainage; DISTLINEA is distance from lineament; ASPECTCw, CURVAw; TOPOw, TEXTUREw, DRAINw, MATERIALw, THICKw, TYPEw, DIAMETERw, AGEw; DENSITYw, GEOLw and LANDUSEw are logistic regression coefficients; z is a parameter; and p is the landslide-occurrence possibility.

Using these formulae, a landslide susceptibility maps were made. Using formulas (1) and (2), the landslide susceptibility maps were made in the Sagimakri and the other study area, Samgyori. Also, using formulas (1) and (3), the landslide susceptibility maps were made in the Samgyori and the other study area, Sagimakri. To compare the validation result quantitative, the AUC (areas under the curve) method was used. [2] In the case of Sagimakri based on logistic regression formula and coefficient of Sagimakri, the prediction accuracy is 87.07%. In the case of Sagimakri based on logistic regression formula and coefficient of Samgyori, the prediction accuracy is 73.85%. In the case of Samgyori based on logistic regression formula and coefficient of Samgyori, the prediction accuracy is 92.18%. In the case of Samgyori based on logistic regression formula and coefficient of Sagimakri the prediction accuracy is 80.66%.

4. DISCUSSION AND CONCLUSION

In this study, a cross application and validation of statistical approach to estimating the susceptible area of landslides using GIS is presented. With respect to the validation and cross-validation of Samgyori study area, the success rates of the logistic regression method showed more accurate result than the rates of Sagimakri study area. A study area reflected on the basis of its own logistic regression formula and coefficient showed accuracy greater than that based on alternative area data. Generally, the validation results showed satisfactory agreement between the susceptibility map and the existing data on landslide location. Landslides are among the most hazardous natural disasters. So, government and research institutions worldwide have attempted for years to assess the landslide hazard and risk and to show its spatial distribution. The Landslide susceptibility maps are of great help to planners and engineers in their choosing of areas for further detail survey and of locations suitable for development.

5. REFERENCES

[1] Atkinson, P.M. and Massari, R., Generalized linear modeling of susceptibility to landsliding in the central Apennines, Italy. *Computer & Geosciences* 24, 373-385, 1998

[2] Lee, S. and Sambath, T., Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental Geology* 50, 847-855, 2006

[3] Lee, S., Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *International Journal of Remote Sensing* 26, 1477-1491, 2005

[4] Ohlmacher, G.C. and Davis, J.C., Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansa, USA. *Engineering Geology* 69, 331-343, 2003