Application of a weight-of-evidence model to landslide susceptibility analysis Boeun, Korea

Moungjin Lee[1)*], Youngtae Yu[2)]

1) Department of Earth System Sciences, Yonsei University, 134 Shinchon-dong Seodaemun-gu, Seoul 120-749, Korea Tel: 82-2-2123-2673; E-mail address: master2003@yonsei.ac.kr

2) Department of Computer Engineering, Hannam University, 133 Ojung-dong Daeduk-gu, Taejeon 306-010, Korea, Tel: 82-42-868-3057; Fax: 82-42-867-0421; E-mail address: ytyu@kis.kigam.re.kr

Abstract : The weight-of-evidence model one of the Bayesian probability model was applied to the task of evaluating landslide susceptibility using GIS. Using the location of the landslides and spatial database such as topography, soil, forest, geology, land use and lineament, the weight-of-evidence model was applied to calculate each factor' s rating at Boun area in Korea where suffered substantial landslide damage following heavy rain in 1998. The factors are slope, aspect and curvature from the topographic database, soil texture, soil material, soil drainage, soil effective thickness, and topographic type from the soil database, forest type, timber diameter, timber age and forest density from the forest map, lithology from the geological database, land use from Landsat TM satellite image and lineament from IRS satellite image. Tests of conditional independence were performed for the selection of the factors, allowing the 43 combinations of factors to be analyzed. For the analysis, the contrast value, W^+ and W^-, as each factor' s rating, were overlaid to map landslide susceptibility. The results of the analysis were validated using the observed landslide locations, and among the combinations, the combination of slope, curvature, topographic, timber diameter, geology and lineament show the best results. The results can be used for hazard prevention and planning land use and construction.

INTRODUCTION

The Boun area of study had much landslide damage following heavy rain in 1998 and was selected as a suitable case to evaluate the frequency and distribution of landslides (figure 1). The site lies between the latitudes 36 °25' 21' ' N and 36° 30' 00' ' N, and longitudes 127° 39' 36' ' E and 127° 45' 00' ' E, and covers an area of 68.43km^2. The bedrock geology of the study area consists mainly of biotite granite. In the study area, the landslides were mainly soil slide that occurred and the landslides occurred where the maximum daily rainfall is 407 mm.

DATA AND METHODOLOGY

Maps relevant to landslide occurrence were constructed to a vector type spatial database using the GIS software ARC/INFO. These included 1:5,000 scale topographic maps. The digital elevation model (DEM) was created using the topographic database. 1:25,000 scale soil maps, 1:25,000 scale forest maps and 1:50,000 scale geological maps. Slope, aspect and curvature, which are relevant to the landslide analysis, were calculated from the DEM. Soil texture, soil material, soil drainage, soil effective thickness, and topographic type were extracted from the soil database. Forest type, timber diameter, timber age, and forest density were extracted from the forest map. Lithologies were extracted from the geological database. Land use and lineament data were extracted from satellite image. The land use data was classified from Landsat TM image with resolution 30m × 30m. The lineament was detected from IRS image with resolution 5m × 5m by expert of structural geologist and distance from lineament was calculated by buffering. Using the detected landslide locations and the constructed spatial database, a landslide analysis methods were applied and validated. For this, the calculated and extracted factors were converted to a 5m × 5m grid (ARC/INFO GRID type). The spatial relationships were used as each factor' s rating in the overlay analysis. Then, a tests of conditional independence were performed for the selection of the factors to be used for the landslide susceptibility mapping.

WEIGHT OF EVIDENCE MODEL

The following formulation of the Bayesian probability model, known as the weight-of-evidence model, was applied to landslide susceptibility analysis as synthesized from Bonham-Carter and others (1989), Bonham-Carter (1994), and Emmanuel and others (2000). For a given number of units cells, $N\{D\}$, containing an occurrence, D, (figure 2), the prior probability of an occurrence is expressed by

$$P\{D\} = \frac{N\{D\}}{N\{T\}} \quad (1)$$

Now suppose that a binary predictor pattern, B, occupying $N\{B\}$ unit cells, occurs in the region, and that a number of known landslides occurs preferentially within the pattern, i.e., $N\{D \cap B\}$, then the favorability of locating an occurrence given the presence of a predictor and the absence of a pattern can be expressed by the conditional probabilities

$$P\{D \mid B\} = \frac{P\{D \cap B\}}{P\{B\}} = P\{D\}\frac{P\{B \mid D\}}{P\{B\}} \quad (2)$$

$$P\{D \mid \bar{B}\} = \frac{P\{D \cap \bar{B}\}}{P\{\bar{B}\}} = P\{D\}\frac{P\{\bar{B} \mid D\}}{P\{\bar{B}\}} \quad (3)$$

The posterior probability of an occurrence given the presence and absence of the predictor pattern are denoted by $P\{D \mid B\}$ and $P\{D \mid \bar{B}\}$, respectively; and $P\{B \mid D\}$ and $P\{\bar{B} \mid D\}$ are the posterior probabilities of being inside and outside the predictor pattern B, respectively, given the presence of an occurrence D. Also, $P\{B\}$ and $P\{\bar{B}\}$ are the prior probabilities of being inside and outside the predictor pattern. The same model can be expressed in an odds-type formulation, where the odds, O, are defined as $O = P/(1-P)$. Expressed as odds, equation (2) and (3), respectively, become

$$O\{D \mid B\} = O\{D\}\frac{P\{B \mid D\}}{P\{B \mid \overline{D}\}} \quad (4)$$

$$O\{D \mid \overline{B}\} = O\{D\}\frac{P\{\overline{B} \mid D\}}{P\{\overline{B} \mid \overline{D}\}} \quad (5)$$

where $O\{D \mid B\}$ and $O\{D \mid \overline{B}\}$are the posterior odds of an occurrence given the presence and absence of a binary predictor pattern, respectively, and $O\{D\}$ is the prior odds of an occurrence. The weights for the binary predictor pattern are defined as

$$W^{+} = \log_e \frac{P\{B \mid D\}}{P\{B \mid \overline{D}\}} \quad (6)$$

$$W^{-} = \log_e \frac{P\{\overline{B} \mid D\}}{P\{\overline{B} \mid \overline{D}\}} \quad (7)$$

where W^{+} and W^{-} are the weights of evidence when a binary predictor pattern is present and absent, respectively. Hence,

$$\log_e O\{D \mid B\} = \log_e O\{D\} + W^{-} \quad (8)$$

$$\log_e O\{D \mid \overline{B}\} = \log_e O\{D\} + W^{-} \quad (9)$$

Now suppose that there are two binary predictor patterns, $B1$ and $B2$. It can be shown that the posterior probability of an occurrence given the presence of the two predictor patterns is

$$P\{D \mid B_1 \cap B_2\} = \frac{P\{B_1 \cap B_2 \mid D\}P\{D\}}{P\{B_1 \cap B_2 \mid D\}P\{D\} + P\{B_1 \cap B_2 \mid \overline{D}\}P\{\overline{D}\}} \quad (10)$$

If $B1$ and $B2$ are conditionally independent of each other with respect to a set of occurrences, then it indicates that the following relation is satisfied

$$P\{B_1 \cap B_2 \mid D\} = P\{B_1 \mid D\}P\{B_1\} \quad (11)$$

This allows equation (10) to be simplified as

$$P\{D \mid B_1 \cap B_2\} = P\{D\}\frac{P\{B_1 \mid D\}P\{B_2 \mid D\}}{P\{B_1\}P\{B_2\}} \quad (12)$$

Similarly, if more than two binary predictor patterns are used, they can be added, provided that they are also conditionally independent of one another with respect to the occurrences. Thus, with Bj (j = 1, 2, ... , n) binary predictor patterns, the log$e$ posterior odds are

$$\log_e O\{D \mid B_1^k \cap B_2^k \cap B_3^k \ldots B_n^k\} = \sum_{j=1}^{n} W_j^k + \log_e O\{D\} \quad (13)$$

where the superscript k is positive (+) or negative (-) depending on whether the binary predictor pattern is present or absent, respectively. The posterior odds can be converted to posterior probabilities, based on the relation $P = O/(1+O)$, to represent the favorability of locating an occurrence.

equation (13) defines the relationship for two binary patterns, $B1$ and $B2$, that are conditionally independent of each other with respect to a set of points. Algebraic manipulation shows that equation (13) is equivalent to

$$N\{B_1 \cap B_2 \cap D\} = \frac{N\{B_1 \cap D\}N\{B_2 \cap D\}}{N\{D\}} \quad (14)$$

The left side of equation (14) is the observed number of occurrences in the overlap zone of $B1$ and $B2$. The right side is the predicted number of occurrences in this overlap zone. This relationship leads to a contingency table calculation for the pair-wise testing of conditional independence.

DATA INTEGRATION AND VALIDATION FOR

LANDSLIDE SUSCEPTIBILITY MAPPING

Then contrast of each factor' s type or range were summed to calculate the landslide susceptibility index, as shown in equation (15)

$$LSIc = \Sigma Fc \text{ (where } Fc = \text{contrast of each factor' s range or type)} \quad (15)$$

Also, the binary predictor patterns were assigned the weights and were integrated according to equation (16).

$$LSIw = \Sigma Fw \text{ (where } Fw = W^{+} \text{ and } W^{-} \text{ of binary patter of each factor' s range or type)} \quad (16)$$

To generate the binary predictor pattern of the 15 factors, the spatial database was reclassified into a binary pattern as ' ' favorable' ' and the other formations as ' ' nonfavorable.' ' To generate the binary predictor patterns of each factors, we have to determine the rating or range for which the spatial association between the landslide occurrences and each factors is optimal. The optimum cutoff for binary pattern is determined by calculating the C/S(C), studentized value of contrast. Then, the W^{+} and W^{-} value used as ratings of each factors.

If the LSIc and LSIw value is high, it means a higher susceptibility to landslide; a lower value means a lower susceptibility to landslides. Before the integration, the statistical validity of the resulting predictive maps is examined by applying an overall test of conditional independence. All of the factors and the 43 combinations of the factors that were determined from the tests of conditional independence of the contrast were used for the landslide susceptibility analysis.

The success rates illustrate how well the estimators perform with respect to the landslides used in constructing those estimators (Chung and Fabbri, 1999). The range was classified by area, which was equal to about 5%. For example, in Case All, an index value above 90 that is 10%, could be explained by 49% of the landslides, and in Case S1, an index value above 90 could explained 47% of the landslides. Also, in Case All, an index value above 70 that is 75%, could explained 60% of the landslides. and in Case S1, an index value above 70 could explained 79% of the landslides. Among the 42 cases considered, Case S15 showed the best result, 52% and 81% of landslides for above 90 and 70 in index each. The Case S15 was used slope, aspect, curvature, topograhic type, timber diameter, geology and lineament after the test of conditional independence.

DISCUSSION AND CONCLUSION

this study, a probabilistic, weight-of-evidence model, approach to estimating the susceptible area of landslides using the GIS is presented.

In the Boun area, landslide occurrence locations detected from aerial photograph interpretation and field surveys were formed into a GIS database. Various maps were constructed from the landslide-related factors derived from the database. These include a 1:5,000 scale topographic map, 1:25,000 scale soil maps, 1:25,000 scale forest map and 1:50,000 scale geological map. The land use and lineament detected and classified from satellite images. The factors involved are aspect, slope and curvature as derived from the topography database, soil texture, soil material, soil drainage, soil effective thickness and topographic type as derived from the soil database, forest type, timber diameter, timber age and forest density as derived from the forest database, lithology as derived from the geology database, land use classified from Landsat TM image and lineament detected from IRS image. The relationship of landslide and factors analyzed for landslide hazard assessment using the weight-of-evidence method. Then, a test of independence between each factor was performed, and then 42 combinations of the factors were derived. For the analysis, the calculated contrast, W^+ and W^- value using the weight-of-evidence method, as each factor's rating, were overlaid to calculate a landslide susceptibility index. Then indexes were mapped, and the results were validated by calculating the correlation between the observed landslide location and the results. Among the 42 combinations, S15 that used slope, curvature, topographic type, timber diameter, geology and lineament after the test of conditional independence showed the best prediction accuracy.

Landslide susceptibility maps are very helpful to planners and engineers for choosing suitable locations to carry out developments. However, 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. For the method to be applied in general, more landslide location data are needed, as well as its application to more regions. Fortunately, the landslide-related spatial database for topography, soil, forest, and geology is already available for most areas of Korea, so the landslide analysis can be performed quickly and cheaply for all of Korea.

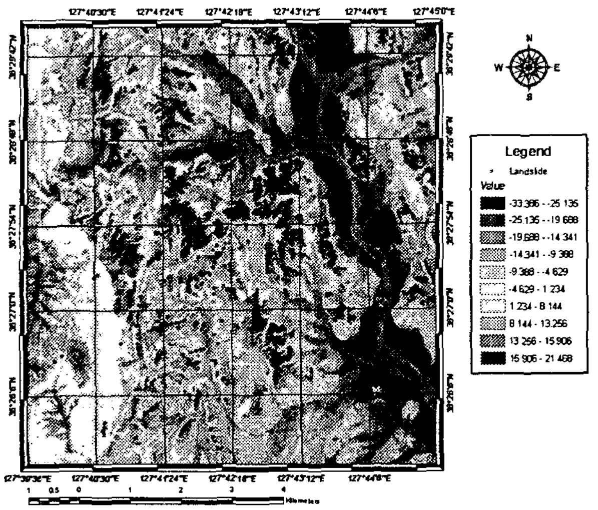

Figure 1. Predictive model map of landslide based on weights of evidence analysis; using all contrasts.

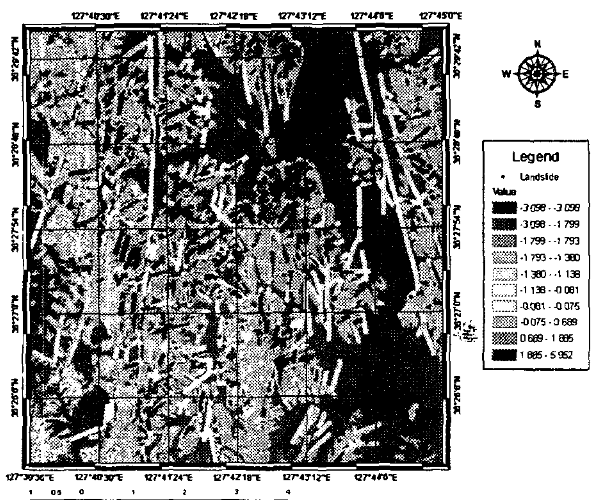

Figure 2. Predictive model map of landslide based on weights of evidence analysis; using slope, wood diameter, curvature, topography, geology and lineament after the test of conditional independence.

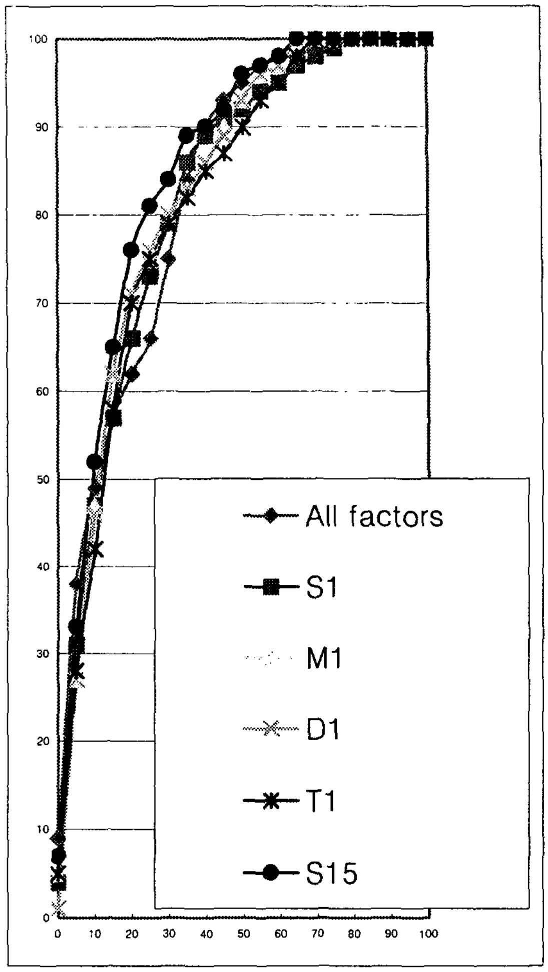

Figure 3. Illustration of cumulative frequency diagram showing landslide susceptibility index rank (y-axis) occurring in cumulative percent of landslide occurrence(x-axis).

REFERENCES

AGTERBERG, F. P., BONHAM-CARTER, G. F., and WRIGHT, D. F., 1990, Statistical pattern integration for mineral exploration. In Computer Applications in Resource Estimation Prediction and Assessment for Metals and Petroleum edit by G., Gaal, and D. F., Merriam (Pergamon Press: Oxford), pp. 1- 21.

ASADI, H. H., and HALE, M. L., 2001, A prddictive GIS model for mapping potential gold and base metal mineralization in Takab area, Iran. Computers & Geosciencess, 27, 901-912,