Probabilistic Landslide Susceptibility Analysis and Verification using GIS and Remote Sensing Data at Penang, Malaysia

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

En. Jasmi Ab Talib
Spatial Data Analysis & Modeling Division, Malaysian Centre For Remote Sensing (MACRES),
No.13, Jalan Tun Ismail, 50480, Kuala Lumpur, Malaysia
jasmi@macres.gov.my

Abstract: The aim of this study is to evaluate the hazard of landslides at Penang, Malaysia, using a Geographic Information System (GIS) and remote sensing. Landslide locations were identified in the study area from interpretation of aerial photographs and field surveys. The topographic and geologic data and satellite image were collected, processed and constructed into a spatial database using GIS and image processing. The used factors that influence landslide occurrence are topographic slope, topographic aspect topographic curvature and distance from drainage from topographic database, geology and distance from lineament from the geologic database, land use from TM satellite image and vegetation index value from SPOT satellite image. Landslide hazardous area were analysed and mapped using the landslide-occurrence factors by probability - likelihood ratio - method. The results of the analysis were verified using the landslide location data. The validation results showed satisfactory agreement between the hazard map and the existing data on landslide location.

Keywords: Key words: Landslide; susceptibility; likelihood ratio; GIS; Remote Sensing; Penang

1. Introduction

Recently there has been an increasing occurrence of landslides in Malaysia. Most of these landslides were on cut slopes or embankments along the road and highways in mountainous area. Some of these landslides occurred near high-rise apartment and residential area, creating great anxiety to various groups of people. A few major and catastrophic landslides also occurred within the last ten years. The landslides include the tragis Highland Tower landslide, Genting Sempah landslide, Gua Tempurung landslide, Paya Terubung landslide and Bukit Antarabangsa landslides. The frequent landslides result in significant damage to people and property. In the study area, Penang in Malaysia, much damage was caused on these occasions. The reason for the landslides was heavy rainfall, and, as there was little effort to assess or predict the event, damage was extensive. Through scientific analysis of landslides, we can assess and predict landslide-susceptible areas, and thus decrease landslide damage through proper preparation. In order to achieve this, landslide hazard analysis techniques were developed, applied, and verified in the study area. GIS (Geographic Information System) software, ArcView 3.2 and ARC/INFO 8.1 NT version, was used as the basic analysis tool for spatial management and data manipulation.

The Penang area of study had much landslide damage following heavy rain and was selected as a suitable case to evaluate the frequency and distribution of landslides. Penang is one of the 13 states of the Federation of Malaysia. The Penang is located on the northwest coast of Peninsula Malaysia. It is bounded on the north and east by the state of Kedah, to the South by the State of Perak and to the west by the Straits of Malacca and Sumatra (Indonesia). Penang consists of the island of Penang and a coastal strip on the mainland known as Province Wellesley. The island covering an area of $285km^2$ is separated from the mainland by a channel. The rainfall is quite evenly distributed throughout the year, with more rain from September to November. The whole state of Penang has a population of approximately 1 million people. The bedrock geology of the study area consists mainly of granite.

For the landslide-hazard analysis, we have collected data, made spatial database, extracted landslide related factors, calculated likelihood ratio, overlaid the factors, made landslide susceptibility map and verified the map. A key assumption using probability, likelihood ratio, approach is that the potential (occurrence possibility) of landslides will be comparable to the actual frequency of landslides. Landslide occurrence areas were detected in the Penang area, Malaysia by interpretation of aerial photographs and field surveys. A map of landslides was developed from aerial photographs, in combination with the GIS, and this were used to evaluate the frequency and distribution of shallow landslides in the area. Topography and lithology databases were constructed and lineament, land use and vegetation index value extracted from Landsat TM and SPOT XS satellite image for the analysis. Maps relevant to landslide occurrence were constructed to a vector type spatial database using the GIS software ARC/INFO. Using the database, landslide related factors were extracted. First, using the topographic database, the digital elevation model (DEM) was created and using the DEM, slope, aspect and curvature. In addition, using the topographic database, the distance from drainage was calculated. Using the geology database, lithology is extracted and distance from lineament was calculated. The land use was extracted from Landsat TM satellite image and field checks and vegetation index value calculated from SPOT satellite image.

For this, the calculated and extracted factors were converted to a 10m × 10m grid (ARC/INFO GRID type). Then, using univariant probability analysis, likelihood ratio method, the spatial relationships between the landslide location and each landslide-related factor, were analyzed. The relationship was used as each factor's rating in the overlay analysis. So, the factor's ratings were summed to landslide hazard index and hazard mapping. Finally, the hazard map was verified using existing landslide location.

2. Data using GIS and Remote Sensing

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. In this study, 1:10,000 –1:50,000 scale aerial photographs used to detect landslide locations. The photographs were taken from 1981 to 2000. The landslide locations detected by photo interpretation and the locations were verified by fieldwork. Recent landslides were observed in aerial photographs as a break in the forest canopy, bare soil, or other geomorphic characteristics typical of landslide scars; for example, head and side scarps, flow tracks, and soil and debris deposits below the scar. To assemble a database to assess the surface area and number of landslides in each of three study areas. In total, 541 landslides were mapped in a total area of 293km$^{2.}$

There are 8 factors considered in calculating the probability. The factors were extracted from the constructed spatial database. Contour and survey base points that have an elevation value read from the 1:50,000-scale topographic map were extracted, and a DEM (Digital Elevation Model) was made. The DEM has the 10m resolutions. Using the DEM, the slope angle, slope aspect and slope curvature were calculated. Distance from drainage was calculated in 100m intervals. The lithology map was obtained from a 1:50,000-scale geologic map. The distance from linearment was calculated in 100m intervals. Land use data classified from a LANDSAT TM image using unsupervised classification method and field survey. The 11 classes such as urban, water, forest, agricultural area and barren area were extracted for land use mapping. The land use was classified from Landsat TM satellite image. Finally, vegetation index value (NDVI) was calculated from SPOT satellite image. The Normalized Difference Vegetation Index (NDVI) is calculated using the formula for NDVI is that (IR – R) / (IR + R), where IR stands for the infrared portion of the electromagnetic spectrum, and R stands for the red portion of the electromagnetic spectrum. The NDVI finds areas of vegetation in imagery.

3. Application of probabilistic model

Generally, for predicting the landslide, the following assumption is necessary. Landslide occurrence is determinated from landslide related factor and the future landslide can be occurred in the same condition with past landslide. On the base of the assumption, the relationship between landslide occurred area and landslide related factors can be distinguish from the relationship between landslide not occurred area and landslide related factors. To represent the distinction quantitatively, the likelihood ratio was used for this study. The likelihood ratio is ratio of probability that is occurrence probability to not-occurrence probability in certain attribute (Bonham-Carter, 1994). In the case of landslide, if we set the landslide occurrence event to B and certain factors' attribute to D, the likelihood ratio in D is ratio of conditional probability. Therefore, the ratio is higher than 1, the higher relationship between landslide and the certain factors' attribute and the ratio is lower than 1, the lower relationship between landslide and the certain factors' attribute. For calculating of the likelihood ratio, the table was made for each landslide related factors. Then area ratio for landslide occurrence and not occurrence was calculated for range or type of each factor, and area ratio for range or type of each factor to total area was calculated. Finally, likelihood ratios for range or type of each factor was calculated by divide the landslide occurrence ratio by the area ratio.

4. Landslide susceptibility mapping and verification

Using the probability method, the spatial relationship between landslide-occurrence location and each landslide-related factor was derived. The used factors that influence landslide occurrence are topographic slope, topographic aspect topographic curvature and distance from drainage from topographic database, geology and distance from lineament from the geologic database, land use from TM satellite image and vegetation index value from SPOT XS satellite image. The factors were converted to a 10 m × 10 m grid for calculate the landslide hazard index. In the study area, the total cell number is 2,928,378 and the landslide occurrence cell number is 541 Using GIS software, a grid was overlain with each geographic coverage for each study area.

The correlation ratings are calculated from relation analysis between landslides and the relevant factors. Therefore, the rating of each factor's type or range was assigned as the relationship between landslide and each factor's type or range, that is, ratio of the number of cells where landslides not occurred to the number of cells where landslides occurred. The landslide hazard index (LHI) is calculated by summation of each factor's ratio value.

$$LHI = \Sigma Fr \text{ (Fr: Rating of each factors' type or range)} \quad (1)$$

After the calculation, the LHI have minimum value, 1.93 and maximum value, 15.82. The average value is 8.02 and standard deviation is 2.32. The relation analysis is the ratio of the area where landslides occurred to the total area, so a value of 8 means an average value because 8 factors used. If the value is greater than 8, it means a higher correlation, and lower than 8 means lower correlation. The landslide-hazard map was made using the LHI value index for interpretation. The index is classified by equal areas and grouped into five classes.

For verification of landslide hazard calculation methods, two basic assumptions are needed. One is that landslides are related to factors such as slope, aspect, curvature, distance from drainage, geology, distance from lineament, land use and vegetation index, and the other is that future landslides will be predicted by a specific impact factor such as rainfall or earthquake. In this study, the two assumptions are satisfied because the landslides are related to the spatial information and the landslides were occurred by one cause, heavy rainfall in the study area.

The success rate verification results, from comparing the hazard calculation results and landslide occurrence location using likelihood method, appear as a line graph. The success rate illustrates how well the estimators perform (Chung and Fabbri, 1999). To obtain the relative ranks for each prediction pattern, the calculated index values of all cells in the study area were sorted in descending order. Then the ordered cell values were divided into 100 classes, with accumulated 1% intervals. The above procedure also was adapted for the right side of the study area by comparing the 100 classes obtained with the distribution on the study area. As a result, an index value above 11.33 that is 10% of the study area where landslide susceptibility index is higher in rank could be explained by 41% of the landslides of the all landslides. In addition, an index value above 10.60 that is 30% could be explained by 68% of the landslides.

5. Discussion and conclusion

Landslides are among the most hazardous natural disasters. Government and research institutions worldwide have attempted for years to assess the landslide hazard and risk and to show its spatial distribution. In this study, a probabilistic approach to estimating the susceptible area of landslides using GIS and remote sensing is presented. In this study, only the hazard analysis was performed, because the small area studied did not allow us to determine the distribution of rainfall. However, if data on factors causing the landslides, such as rainfall, earthquake shaking, or slope cutting, exist, then the possibility analysis could also be done. If the factors relevant to vulnerability of buildings and other property were available, risk analysis could also be done. Landslide hazard 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 the 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 more generally applied, more landslide data are needed, as well as application to more regions.

References

[1] Bonham-Carter GF. 1994. *Geographic Information Systems for geoscientists, modeling with GIS*. Oxford: Pergamon Press.

[2] Chung CF, Fabbri AG.1999. Probabilistic prediction models for landslide hazard mapping, *Photogrammetric Engineering & Remote Sensing* 65:1389-1399.