PROBABILISTIC LANDSLIDE SUSCEPTIBILITY AND FACTOR EFFECT ANALYSIS

[1]SARO. LEE and JASMI , AB TALIB[2]
[1]Geoscience Information Center, Korea Institute of Geology & Mineral Resources (KIGAM)
30, Gajung-dong, Yusung-gu, Daejeon, 305-350, Korea
[2]Spatial Data Analysis & Modeling Division, Malaysian Centre For Remote Sensing (MACRES)
leesaro@kigam.re.kr

Abstract: The susceptibility of landslides and the effect of landslide-related factors at Penang in Malaysia using the Geographic Information System (GIS) and remote sensing data have been evaluated. Landslide locations were identified in the study area from interpretation of aerial photographs and from field surveys. Topographical and geological data and satellite images were collected, processed, and constructed into a spatial database using GIS and image processing. The factors chosen that influence landslide occurrence were: topographic slope, topographic aspect, topographic curvature and distance from drainage, all from the topographic database; lithology and distance from lineament, taken from the geologic database; landuse from Landsat TM (Thermatic Mapper) satellite images; and the vegetation index value from SPOT HRV (High Resolution Visible) satellite images. Landslide hazardous areas were analysed and mapped using the landslide-occurrence factors employing the probability–frequency ratio method. To assess the effect of these factors, each factor was excluded from the analysis, and its effect verified using the landslide location data. As a result, land cover had relatively positive effects, and lithology had relatively negative effects on the landslide susceptibility maps in the study area. In addition, the landslide susceptibility maps using the all factors showed the relatively good results.

Keywords: Landslide; Frequency ratio; Effect analysis; GIS; Remote Sensing; Penang; Malaysia

1. INTRODUCTION

Recently there has been an increasing occurrence of landslides in Malaysia. Most of these landslides occurred on cut slopes or on embankments alongside roads and highways in mountainous areas. Some of these landslides occurred near high-rise apartments and in residential areas, causing great anxiety in many people. A few major and catastrophic landslides have also occurred within the last 10 years. These landslides have resulted in significant damage to people and property. In the area chosen in this study, Penang in Malaysia, much damage was caused on each of these occasions. The trigger for the landslides was a period of heavy rainfall, and, as there was little effort to assess or predict the event, damage was extensive. Through scientific analysis of landslides, landslide-susceptible areas can be we can be assessed and predicted and thus landslide damage can be decreased through proper preparation. To achieve this aim, landslide susceptibility analysis techniques have been applied, and verified in the study area. In addition, landslide-related factors were also assessed. Geographic Information System (GIS) software, ArcView 3.2, and ARC/INFO 8.1 NT version software packages were used as the basic analysis tools for spatial management and data manipulation.

The Penang area has suffered much landslide damage following heavy rains, and was selected as a suitable candidate to evaluate the frequency and distribution of landslides. Penang is one of the 13 states of the Federation of Malaysia. The Penang area is the northwest coast of the Malaysian peninsula. It is bounded to 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 by Sumatra (Indonesia). Penang consists of the island of Penang, and a coastal strip on the mainland, known as Province Wellesley. The island covers an area of 285 km^2, and is separated from the mainland by a channel. The rainfall is quite evenly distributed throughout the year, with more rain occurring from September to November. Penang has a population of approximately one million people. The bedrock geology of the study area consists mainly of granite.

For the analysis of landslide susceptibility and for the assessment of landslide-related factors, we have collected data, made a spatial database, extracted landslide-related factors, calculated frequency ratios, overlaid the factors, made landslide susceptibility maps, and verified the effect of each factor on the maps. A key assumption using the probability–frequency ratio approach is that the potential (occurrence possibility) of landslides is comparable to the actual frequency of landslides. Landslide-occurrence areas were detected in the Penang area by interpreting aerial photographs and field surveys. A map of landslides was developed from aerial photographs, in combination with the GIS data, and this was used to evaluate the frequency and distribution of shallow landslides in the Penang area. For analysis of the landslide-related factors, topography and lithology databases were constructed, and the lineament, land use, and vegetation index values were extracted from Landsat TM and SPOT XS satellite images.

2. Data gathering using GIS and remote sensing images

Accurate detection of the location of landslides is very important for probabilistic landslide susceptibility analysis. The application of remote sensing methods, such as aerial photographs and satellite images, are used to obtain significant and cost-effective information on landslides. In this study, 1:10,000–1:50,000-scale aerial photographs were used to detect the landslide locations. These photographs were taken during the period 1981–2000, and the landslide locations were detected by photo interpretation and the locations verified by fieldwork. Recent landslides were observed in aerial photographs from breaks 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 a scar. To assemble a database to assess the surface area and number of landslides in each of three study areas, a total of 541 landslides were mapped in a mapped area of 293 km^2.

To apply the probabilistic method, a spatial database that considers landslide-related factors was designed and constructed. These data are available in Malaysia either as paper or as digital maps.

There were eight factors considered in calculating the probability, and the factors were extracted from the constructed spatial database. The factors were transformed into a vector-type spatial database using the GIS, and landslide-related factors were extracted using our database. Using the topographic database, a digital elevation model (DEM) was created first. Contour and survey base points that had elevation values read from the 1:50,000-scale topographic maps were extracted, and a DEM was constructed with a resolution of 10 m. Using this DEM, the slope angle, slope aspect, and slope curvature were calculated. In addition, the distance from drainage was calculated using the topographic database. The drainage buffer was calculated in 100 m intervals. Using the geology database, the lithology was extracted, and the distance from lineament calculated. The lithology map has obtained from a 1:50,000-scale geological map. The lineament buffer was calculated in 100 m intervals. Land use data was classified using a LANDSAT TM image employing an unsupervised classification method and field survey. The 11 classes identified, such as urban, water, forest, agricultural area, and barren area were extracted for land use mapping. Finally, the Normalized Difference Vegetation Index (NDVI) has obtained from SPOT satellite images. The NDVI value was calculated using the formula NDVI = (IR − R)/(IR + R), where IR value is the infrared portion of the electromagnetic spectrum, and R-value is the red portion of the electromagnetic spectrum. The NDVI value denotes areas of vegetation in an image.

3. THE FREQUENCY RATIO METHOD AND THE RELATIONSHIP BETWEEN LANDSLIDES AND FACTORS

In general, to predict landslides, it is necessary to assume that landslide occurrence is determined by landslide-related factors, and that future landslides will occur under the same conditions as past landslides. Using this assumption, the relationship between landslides occurring in an area and the landslide-related factors can be distinguished from the relationship between landslides not occurring in an area and the landslide-related factors. To represent this distinction quantitatively, we used the frequency ratio. The frequency ratio is the ratio of the area where landslides occurred to the total study area, and also, is the ratio of the probabilities of a landslide occurrence to a non-occurrence for a given attribute. In the case of landslide occurrence, if the landslide-occurrence event is denoted by "B", and a given factor's attribute is denoted by "D", then the frequency ratio of D is the ratio of the conditional probabilities. Therefore, the greater this ratio is above unity, then the stronger the relationship between landslide occurrence and the given factor's attribute. The lower the ratio is below unity, then the lesser the relationship is between landslide occurrence and the given factor's attribute. To calculate the frequency ratio, a table was constructed for each landslide-related factor. Then, the area ratios for landslide occurrence and non-occurrence were calculated for each range or type of factor, and the area ratio for each range or type of factor to the total area was calculated. Finally, the frequency ratios for each range or type of factor were calculated by dividing the landslide-occurrence ratio by the area ratio.

The factors chosen, such as the slope, aspect, curvature, distance from drainage, geology, distance from lineament, land use, and vegetation index were evaluated using the frequency ratio method to determine the level of correlation between the location of the landslides in the study area and these factors. Probabilistic approaches are based on the observed relationships between each factor and the distribution of landslides.

Topographic factors, such as slope, aspect, curvature, and distance from drainage were used, and the relationship between landslide occurrence and slope, show that steeper slopes have greater landslide probabilities. Below a slope of 5°, the ratio was <1, which indicates a very low probability of landslide occurrence of 0.26. For slopes above 6°, the ratio was >1, which indicates a high probability of landslide occurrence. This means that the landslide probability increases according to slope angle. As the slope angle increases, then the shear stress in the soil or other unconsolidated material generally increases. Gentle slopes are expected to have a low frequency of landslides because of the generally lower shear stresses associated with low gradients. Steep natural slopes resulting from outcropping bedrock, however, may not be susceptible to shallow landslides. In the case of the aspect, landslides were most abundant on south-facing and northeast-facing slopes. The frequency of landslides was lowest on east-facing, west-facing, and northwest-facing slopes, except in flat areas. The curvature values represent the morphology of the topography. A positive curvature indicates that the surface was upwardly convex at that cell. A negative curvature indicates that the surface was upwardly concave at that cell. A value of zero indicates that the surface was flat. According to the curvature results, the more positive or negative a value is, then the higher is the probability of a landslide occurrence. Flat areas had a low curvature value of 0.60. The reason for this is that following heavy rainfall, a convex or concave slope contains more water and retains this water for a longer period. Analysis was carried out to assess the influence of drainage lines on landslide occurrence. For this purpose, the proximity to a drainage line was identified by buffering. It can be seen that as the distance from a drainage line increases, the landslide frequency generally decreases. At a distance of <400 m, the ratio was >1, indicating a high probability of landslide occurrence, and at distances >600 m, the ratio was 0, indicating zero probability. This can be attributed to the fact that terrain modification caused by gully erosion may influence the initiation of landslides.

For geological factors, such as the geology and the distance from a lineament, it was found that in the case of the geology, the frequency ratio was higher in granite areas, at 1.25, and was lower in alluvium areas, at 0.53. In the case of the distance from a lineament, the closer the distance was to a lineament, then the greater was the landslide-occurrence probability. For distances to a lineament of <800 m, the ratio was >1, indicating a high probability of landslide occurrence, and for distances to a lineament of >800 m, the ratio was <1, indicating a low probability landslide occurrence. This means that the landslide probability decreases with increasing distance from a lineament. As the distance from a lineament decreases, the fracture of the rock increases, and in addition, the degree of weathering increases.

Using satellite images such as Landsat TM and SPOT XS images, we extracted landslide-related factors, such as land use and the NDVI. In the case of land use, the landslide-occurrence values were higher in scrub, rubber plantation, and mixed areas, and lower in rice growing, swampy, coconut plantation, barren, and oil palm plantation areas. The reason for this is that landslides occurred mainly in inclined and mountainous areas. In the case of the vegetation index, for NDVI values below 0.20, the frequency ratio was <1, which indicates a low landslide-occurrence probability, and for NDVI values above 0.20, the frequency ratio was >1, indicating a high landslide-occurrence probability. This result means that the landslide probability increases according to the vegetation index value.

4. LANDSLIDE SUSCEPTIBILITY ANALYSIS

Using the frequency ratio method, we were able to derive the spatial relationship between a landslide-occurrence location and each landslide-related factor. The factors used were: the topographic slope, topographic aspect, topographic curvature, and distance from drainage, taken from the topographic database; geology, and distance from the lineament, taken from the geologic database; land use from TM satellite images; and vegetation index value from SPOT XS satellite images. The calculated and extracted factors were converted to a 10 × 10 m^2 grid (ARC/INFO GRID-type). In the study area, the total cell number was 2,928,378, and the landslide-occurrence cell number was 541. Using GIS software, a grid was overlain with each geographic coverage for each study area. Then, using univariant probability analysis and the frequency ratio method, the spatial relationship between the landslide location and each landslide-related factor was analysed. The correlation ratings were calculated from analysis of the relationship between the landslides and the relevant factors. Therefore, the rating of each factor's type or range was assigned from the relationship between a landslide and each factor's type or range, that is, the ratio of the number of cells where landslides did not occur to the number of cells where landslides had occurred. The relationship was used to determine each factor's rating in the overlay analysis. A factor's ratings were summed to form the landslide susceptibility index and susceptibility map. The landslide susceptibility index (LSI) was calculated by summation of each factor's ratio value using Equation (1):

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

where Fr is the rating of each factor's type or range.

After calculations using Equation (1), in the case where the slope was excluded, the LSI had a minimum value of 1.99, and a maximum value of 14.12, with an average value of 7.14 and a standard deviation of 1.84.

5. VERIFICATION BY EFFECT ANALYSIS

Effect analysis studies how a solution changes when the input factors are changed, and is a procedure that can determine the effect that a change in a factor can have on the outcome. If the selected factor results in a relatively large change in the outcome, then the outcomes is said to be effective to that factor. This may mean that the factor has to be selected. Effect analysis quantifies the uncertainty of a model to the uncertainty of each factor. The factors that have the greatest impact on the calculated landslide susceptibility map can therefore be identified using effect analysis.

In this work, the effect analyses were conducted by exclusion of each factor in turn during the summation stage using Eq. (1), and the effect of each factor was evaluated. That is, the susceptibility maps using Eq. (1) were verified using an existing landslide location. All eight factors were used, and the LSI values of nine cases including all factors were calculated. For the verification, the method was subjected to tests to determine whether its predictions matched the expected results based on knowledge of the factors, i.e., the authors carried out an effect

analysis in which the model system was subjected to various selections of factors, and the outputs were compared with expected changes in the outputs. Rate curves were created to achieve this. To obtain the rate curves, the calculated landslide susceptibility 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 rate curves explain how well the method and factors predict landslides. To compare the results, the areas below the curves were calculated and re-calculated. A total area = 1 denotes perfect prediction accuracy for all cases. The area below a curve can be used to assess the prediction accuracy qualitatively. For example, in the case where the slope was excluded, 10% of the study area where the landslide susceptibility index had a higher rank could explain 23% of all the landslides. In addition, 30% of the study area where the landslide susceptibility index had a higher rank could explain 55% of the landslides.
From the verification of the model by effect analysis, the land cover, aspect, slope, distance from lineament, distance form drainage, NDVI and curvature (areas below the curve of 0.653, 0.679, 0.681and 0.683, 0.683, 0.686 and 0.690 respectively) have a positive effect (influence) on the landslide susceptibility map using all the factors (area below curve = 0.718). In contrast, lithology, (areas below the curve of 0.721) had a negative influence on the landslide susceptibility map using all factors (area below curve = 0.718). This is because the lower the value of the area below the curve, the greater effect the factor has on the landslide susceptibility map. All factors, except lithology, have a positive effect on the landslide susceptibility map. In contrast, because a higher area below the curve means a more negative effect of the factor on the landslide susceptibility map, factors, only lithology, have a negative effect on the landslide susceptibility map. Because the lithology map of study area is very simple, the lithology can't have positive effect to landslide susceptibility mapping.

6. CONCLUSIONS AND DISCUSSION

Landslides are among the most hazardous of natural disasters. Government and research institutions worldwide have attempted for years to assess landslide hazards and risks, and to show their spatial distribution. Landslide susceptibility maps have been constructed using the relationship between each landslide and causal factors. In this study, a probabilistic approach to estimate susceptible areas to landslides using GIS and remote sensing is presented. Moreover, using effect analysis, the influence of factors on the landslide susceptibility map can be known qualitatively, and the selection of positive factors can improve the prediction accuracy of the landslide susceptibility map. This means that the selection of factors is important to landslide susceptibility mapping. The ratio value from the effect analysis can be used to weight the relative importance of these factors, and can improve the prediction accuracy of the landslide susceptibility map.
In this study, only susceptibility analysis was performed, because the small area studied did not allow the determination of the distribution of any rainfall. However, if data on factors causing the landslides, such as rainfall, earthquakes, or slope cutting exist, then possibility analysis can also be carried out. If the factors relevant to the vulnerability of buildings and other property are available, then risk analysis on this can also be carried out. These results can be used as basic data to assist slope management and land-use planning, but the methods used in this study are also valid for generalized planning and assessment purposes, although they may be less useful on 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.