A GIS-based landslide hazard assessment by multivariate analysis

Xiaoduo PAN\(^{a,b}\), Hiroyuki NAKAMURA\(^{b}\), Tamotsu NOZAKI\(^{c,d}\) and Xiaozhong HUANG\(^{d}\)

Abstract

The occurrence of landslides generally depends on complex interactions among a large number of partially interrelated factors. It is appropriate to use multivariate analysis for predicting landslides from a given set of independent variables. However, the procedure of landslide hazard assessment by regression analysis requires the evaluation of the spatially varying terrain conditions as well as spatial representation of the landslides. In this paper, multivariate regression analysis was applied to predict landslides in the Himi district from independent factors, such as geology, clinial-aspect, slope angle, land use and soil using Geographic Information System (GIS). Based on GIS, every factor was classified into several categories and then the statistical weight of every cluster was assigned for every factor respectively. By the weights of five factors, the linear regression's coefficients of these input factors in landslide area were extracted and assigned to the whole region, and then the susceptibility for the potential landslide was obtained to make the landslide hazard assessment map. According to the coefficients in the Himi district, geology and clinial-aspect factors are the most important ones. Soil factor is not so notable in this research region, though it may be significant in other regions. Finally, the average susceptibilities map for existing landslides was made for the engineers to do control work.

Key words: landslide hazard assessment, GIS, multivariate analysis, coverage, Digital Terrain Model (DTM)

1. Introduction

Landslides have caused untold numbers of casualties and huge economic loss. In many countries, economic loss due to landslides is great and is growing as development expands into unstable hillside areas under the pressures of expanding population. In addition to killing people and animals (both livestock and wildlife), landslides destroy or damage residential and industrial developments areas as well as agricultural and forested areas and negatively affect water quality in rivers and streams (Turner and Schuster, 1996). Therefore, it is necessary to make landslide hazard assessments to help identify dangerous areas for development by examining the potential risk of landslides. Furthermore, once a landslide hazard prone area is identified, management projects can be developed to avoid, prevent, or substantially mitigate the potential hazard.

The occurrence of slope failure generally depends on complex interactions among a large number of partially interrelated factors. Analysis of the landslide hazard requires an evaluation of the relationship between various terrain conditions and landslide occurrences. It is appropriate to use multivariate analysis for the prediction of landslide phenomena (in the form of susceptibility) based on a given set of independent variables, such as geology, clinial-aspect (relationship between dipping direction of strata and slope), slope angle, land use and soil.

However, the procedure of making landslide hazard assessment by regression analysis requires the evaluation of spatially varying terrain conditions as well as spatial representation of the landslides. A GIS allows the storage and manipulation of information concerning the different terrain factors as distinct data layers and thus provides an excellent tool for landslide hazard assessment (Zaitchik and Van, 2003).

Among the most popular statistical landslide hazard methods reported in the recent literature is logistic regression. Logistic regression relates predictor variables (topographic factors, land use, soils etc.) to the presence or absence of landslides (Dai et al., 2001; Ohmacher and Davis, 2003; Dai and Lee, 2003; Santacana et al., 2003; Ayalew et al., 2005), but in recent studies, the fatal weak is that the spatial data should be exported to the statistical software to carry out regression analysis.

An artificial neural network (ANN) offers a computational mechanism that is able to acquire, represent, and compute a mapping from one multivariate space of information to another, given a set of data representing the relationships (Lu and Rosenbaum, 2003). An ANN is trained by the use of a set of associated input and output values. The method is not available within existing GIS systems, and has been programmed in systems like MATLAB (Lee et al., 2003).

In this paper, the convenient tool-GIS was used through the whole process, not only to get the statisti-
cal weights, but also to do the regression analysis.

The objective of this paper is to make landslide hazard assessments by multiple linear regression analysis using ArcGIS 9.0, considering not only the terrain-related variables such as slope angle and land use, but also subterranean variables such as geology, climatic aspect and soil.

2. Landslide distribution and geological setting in the Himi District

The Himi District of The Toyama Prefecture (Fig. 1, (136°52'30" E, 36°55'N), (137°3'E, 36°55'37"N)) in Japan was selected as the research region because hundreds of landslides have occurred in its 145km² area. The northwestern rim of the study area is the ridge line of the mountain which coincides with the borders between Toyama and Ishikawa prefectures. The elevation of the ridge is about 200-500m, and the whole area has a clear tendency to descend towards the Toyama Bay, which is concordant with the bedding aspect of sedimentary strata. Fig. 1 shows the outline of landslide distribution and geology of the study area. Landslide distribution is based on the landslide distribution map of the Himi District (unpublished), 1 : 25,000 in scale, drawn by one of the authors from the interpretation of aerial photographs. The minimum size of the interpreted landslide is 1.2 ha, which is limited by the scale of photographs (1/20,000). Landslides in this research region were classified to 4 categories: rock slide, weathered rock slide, debris slide, clayey soil slide. They were inferred from the topographical feature of each unit and the information of known units by field survey as mentioned above. As a result of the type classification, however, more than 80% of the whole landslide units in the study area were rock slide and weathered rock slide, and the difference between these two types were not so clearly distinguishable. Generally speaking, debris slide and clayey soil slide are clearly smaller than the rock slide or weathered rock slide, because the former types are usually the secondary or tertiary slide of the latter ones and they make clear concave geographical features against the latter types which prevail on the rather convex slope. The maximum size of one unit of the rock slide which might be the primary landslide is about 60 ha and the average is about 20-30 ha. The length of one unit is usually not so longer than the width or even shorter in places, and it does not exceed 2-3 times. In this Himi District, many of the rock slide and weathered rock slide are translational, which have a tendency of slid-

![Fig.1 Landslide distribution and geology in Himi District](image-url)
ing along the bedding plane of the sedimentary strata and the depth to the sliding surface in not so deep. Kuroki, Kunimi and Ikatanji landslides are well-known as typical of this type. Their maximum depth is about 50 m and the average is 20-30 m depending on the recent textbook “Landslides in Toyama 2003” (Japan Landslide Society).

Geology is based on the 1:50,000 in scale geological maps “Abugashima” (Imai et al., 1967a) and “Ochigata” (Imai et al., 1967b). Bedrock of this district mainly consist of soft Miocene-Pleistocene sedimentary rocks. Table 1 shows the stratigraphy and rock facies, and geological members are grouped into 6 categories for the GIS analysis as mentioned below. Bedding plane of the strata trends northeast and gently dips (10°-20°) southeast. Fig. 2 shows the strike-line map compiled from geological maps described above and our own data. The strike-line map is a figure delineated with curves parallel to the strike of strata, and the interval of lines is usually decided in proportion to the cosecant of the dip of strata. In this study area, however, the dip of strata is comparatively consistent and low angle in the whole area (average is 10°-15°). Therefore, the interval of lines in this strike-line map has no meaning and shows only the aspect of strata.

This strike-line map was drawn based on 1/50,000 geological maps “Abugashima” and “Ochigata” (Imai et al., 1967a, b).

Because no clear influence due to faults to the occurrence of landslide has been reported in this area, it is not examined in this paper.

### 3. Method

#### 3.1 Framework

As shown in Fig. 3, the process includes collecting the original maps, creating vector coverages (including point, line and polygon) by digitizing original maps, generating Digital Terrain Models (DTM), and overlaying these DTMs to execute the multivariate analysis to assess the landslide hazard. All work was carried out in GIS software. It was complicated to make the clinial-aspect DTM which indicates the relationship between bedding aspect and surface aspect: in the landslide zone, it was the bedding aspect DTM minus land-

<table>
<thead>
<tr>
<th>Geological time</th>
<th>Geological formation</th>
<th>Rock facies</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quaternary</td>
<td>St. Okayama Member</td>
<td>Sand and clay</td>
<td></td>
</tr>
<tr>
<td>Terrace deposits</td>
<td>Murakami siltstone member</td>
<td>Siltstone (with Sandstone and tuff)</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>Yabuta Siltstone member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Suginoya Siltstone member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nakagawa sandstone member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Calcaceous sandstone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neogene Tertiary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sugata mudstone member</td>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>Ogabe sandstone member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kuzuba alternating beds of sandstone and mudstone</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nakada tuff member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mio sandstone member</td>
<td></td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Takabatake conglomerate member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nakanami mudstone member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ogawa tuffite member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oodomari tuff member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kunimi mudstone member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kakefuda sandstone member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sekidohsan conglomerate member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anamizu Formation</td>
<td>Andesite lava, tuff - breccia and tuff</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>Funatsu granitic rocks</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hida metamorphic rocks</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 1: Stratigraphy of the Himi District*
slide move direction one; in the other zone, it was the bedding aspect DTM minus the aspect DTM which was derived from the Digital Elevation Model (DEM). The DTM of slope angle factor was derived from DEM. The other factors were generated directly using a polygon-grid method.

3.2 Data preparation

The conditioning factors used for landslide susceptibility model based on GIS are well described in literature (Lee et al., 2001; Dai et al., 2001; Clerici et al., 2002; Ohlmercher and Davis, 2003; Lee et al., 2003; Ayalew et al., 2004; Dominguez-Cuesta et al., 2007). However, there is few forecast work done for the Hini District, Toyama Pref, further more, susceptibility assessment was never done for this region before, though landslides happen frequently in this region. Therefore, as many as factors were selected to examine the relationship with landslide according to previous literature, which includes geology, clinal-aspect, slope angle, land use, soil and other factors, such as river density, tree types, and road buffers etc. The factors of geology, clinal-aspect, slope angle, land use and soil were found to have strong relationship with the existing landslides seen from statistical analysis (detail info about statistical analysis is in 3.4 of this paper), so, only these factors were included in regression analysis finally. The data used in this study include the landslide distribution map (including the slide direction), geological maps and the strike-line map above mentioned, topographical maps, 1:25,000 in scale, issued by Geographical Survey Institute, a 1/50,000 land use map (Matsuito, et al., 1984) and a 1/50,000 soil type map (Nogoshi and Uemori, 1984).

Almost all vector coverages were generated by digitizing maps in ArcGIS software and in polygon format, except for the elevation point and arc coverages from the topographic map. After editing, the feature (polygon, line, point) attributes were assigned a numeric value or string. In the landslide distribution coverage, an existing landslide area was assigned the value '1' to identify and a relative landslide moving direction angle. A value of '0' identified other areas. In the strike grid

![Fig. 2 Strike-line map](image)

![Fig. 3 Research framework for this study](image)
-polygon coverage, every polygon of bedrock strata was assigned the dipping direction value. In the elevation arc and point coverages, a related elevation value was assigned to corresponding arc or point. In other polygon coverages (geology, land use and soil), they were named by a relative type name for every polygon.

After digitizing and editing maps, vector coverages were projected with the same projection parameters and converted to raster format. These rasters have the same cell size (25m in this paper) and cover the same geographic extend to allow overlay analysis.

3.3 Classification

There are many different integer values or continuous values in the DTMs from string or numeric coverages. For an example, there are 32 values for types in the soil coverage, for an example, 1 means rock, 2 means lithosols, and so on. It is too detailed to make statistical analyses in this state. Even if the assessment was made on such detailed information, the final map would be very fragmented. Therefore soil types were grouped into categories according to similarity. Other DTMs were also simplified in the following way.

The geology DTM was grouped into 6 categories (see Table 1), other factors’ categories are shown in Table 2.

3.4 Statistical processing

The grouped spatial data cannot be used as input variable values directly in the regression analysis to explore the coefficients between the distribution of landslide and the factors in the study area. The weight \( W_0 \) was taken as input variable values. It was calculated in GIS software by writing DOCELL routines in ArcGIS software and assigned as below:

\[
P_0 = SL_0 + \sum_{i=1}^{m} SL_i \quad (i = 1, 2, \ldots, m)
\]

\[
Q_0 = S_0 + \sum_{i=1}^{n} S_i \quad (i = 1, 2, \ldots, n)
\]

\[
W_0 = P_0 + Q_0 \quad (i = 1, 2, \ldots, m, j = 1, 2, \ldots, n)
\]

Where, \( m \) is the amount of factors, \( m = 5 \) in this research work;

\( n \) is the category amount of every factor DTM, its value depends on every factor’s classification. For an example, there have 6 categories in the geology DTM after the classification, the value of \( m \) is 6;

\( SL_i \) is the area of \( j^{th} \) type in the existing landslide zones for the \( i^{th} \) factor;

\( S_0 \) is the area of \( j^{th} \) type in the whole research area for the \( i^{th} \) factor;

\( P_0 \) is the area-cover percentage of the \( j^{th} \) type in the existing landslide zones for the \( i^{th} \) factor;

\( Q_0 \) is the area-cover percentage of the \( j^{th} \) type in the whole research area for the \( i^{th} \) factor.

In traditional method, because field investigations were difficult to cover the whole region, the weights for landslide factors were calculated only in existing landslide zones, which equal to \( P_0 \). However, some factors that are important for landslide were underestimated and others may be exaggerated in traditional calculation method. Therefore, the ratio of \( P_0 \) to \( Q_0 \) (\( W_0 \)) is more reasonable to assess the potential importance of the environmental factors in the whole region. The development of GIS software makes the value of \( Q_0 \) easier to be calculated and make the value of \( W_0 \) can be obtained. It is obvious to find the difference of the factors between the result of \( P_0 \) in deep color and \( Q_0 \) in light one in Fig. 5.

As a result, before doing the regression analysis in the GIS software, the factor weights were standardized by the following equation:

\[
W_i = \frac{W_0}{\max(W_0, j = 1, 2, \ldots, n)} \quad (i = 1, 2, \ldots, m)
\]

3.5 Multivariate analysis

There are many methods to calculate the susceptibility value for the landslide hazard assessment based on the weights, such as fuzzy analysis, bivariate statistical analysis and ANN. The landslide assessment in this study region was made by multivariate analysis to establish the relationships between numerous input factors.
Table 2  Factors features of the Himi District

<table>
<thead>
<tr>
<th>Group</th>
<th>Source of data</th>
<th>Geology</th>
<th>Strike-line map</th>
<th>Slope angle</th>
<th>Land use</th>
<th>Soil type map</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Imai et al., 1967a/b</td>
<td>Geology map</td>
<td>Strike-line map</td>
<td>Topographical map</td>
<td>Land use map</td>
<td>Soil type map</td>
</tr>
<tr>
<td>B</td>
<td>Same as Table 1</td>
<td>Same as Table 1</td>
<td>30° - 30°: cataclinal</td>
<td>0° - 3°</td>
<td>Paddy field</td>
<td>Brown forest soil</td>
</tr>
<tr>
<td>C</td>
<td>Same as Table 1</td>
<td>Same as Table 1</td>
<td>60° - 120° &amp; -60° - 120°: orthoclinal</td>
<td>3° - 8°</td>
<td>Upland field</td>
<td>Wet brown forest soil &amp; gray upland soil &amp; gley upland soil</td>
</tr>
<tr>
<td>D</td>
<td>Same as Table 1</td>
<td>Same as Table 1</td>
<td>120° - 150° &amp; -120° - 140°: plagio - anacinal</td>
<td>8° - 15°</td>
<td>Forest</td>
<td>Gley soil</td>
</tr>
<tr>
<td>E</td>
<td>Same as Table 1</td>
<td>Same as Table 1</td>
<td>150° - 150°: anacinal</td>
<td>15° - 20°</td>
<td>Bamboo</td>
<td>Dry brown forest soil</td>
</tr>
<tr>
<td>F</td>
<td>Same as Table 1</td>
<td>Same as Table 1</td>
<td>None</td>
<td>30° - 40°</td>
<td>Other use</td>
<td>None</td>
</tr>
<tr>
<td>G</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>More than 40°</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

Usage: SAMPLE (\{mask_grid\}, |grid. _ _ grid|)

\{mask_grid\} - the grid defines the cells to be sampled.
|grid. ..., _grid| - the name of one or more grids whose values will be sampled based on the mask grid. As far as this research work is concerned, mask_grid is the landslide distribution grid file, |grid. ..., _grid| are set of factor grid files related with landslides.

After getting the ASCII file, it is very easy to use the REGRESSION command in GRID mode to calculate the factors coefficients. The model to calculate the coefficients for existing landslide zones is shown in Fig. 6. The resulting coefficients are shown as Table 2. The model to assign the coefficients to calculate the susceptibility for future potential landslides is shown as Fig. 7.

4. Results

After getting the susceptibility values for the whole region, the result can be displayed in ArcView, and can be used as the landslide hazard assessment map shown as Fig. 8.

Using the spatial analysis in ArcView, charts of landslide hazard can be obtained for the whole research area and the existing landslide zones. The chart for the existing landslides area’s susceptibility values is shown in Fig. 9b, and it is clear to see that most of the values are distributed around 1.0. Based on the susceptibilities, the average susceptibility for every existing landslide can be calculated and its rank can be determined as well. The result is shown in Fig. 10.
5. Discussion and conclusions

From Fig. 8, it can be found there is a good agreement between areas of existing landslides and the areas with high susceptibility of landslides. Because of this good agreement, the average probabilities map for existing landslides was made (See index in Fig. 10) and their ranks were determined. The higher the rank is, the more dangerous the according existing landslide is. Therefore, engineers can do control work for existing landslides in the Himi District according to their ranks. Additionally, areas of high probability of a landslide exist outside area of existing landslides. Thus, multivariate regression analysis is applicable to delineate areas of potential landslide hazard.

Table 2 indicates geology factor is the most important factor in determining the susceptibility of landslide occurrence. In general, Hill or Mountain slope of this area gently descends from the northwestern ridge to the Toyama Bay. This topographical aspect is roughly in concordance with the bedding aspect, which means the topography of this area must have been strongly controlled by the geological folding structure. And flexural slip which is a kind of fault along the bedding plane is easy to become a sliding surface of landslide. Some flexural slips were found in this area.

Table 2 also indicates Clinal-aspect factor is more susceptible to landslides. The bed rock in this study region consists of soft sedimentary rocks: sandstone and mudstone or their alternation intercalated with tuff layers in places. These rocks especially mudstone is vulnerable to the weathering and easy to cause slaking repeating wet and dry. Soil factor is not so notable in this study region.

After calculating landslide hazard on the whole region, it can be seen most of the existing landslide values are distributed around 1.0, but the susceptibility of a few areas of the existing landslide-zone is not so high. These areas are thought to be triggered by outer condition, such as snow melting, seismic and spate, however, only the inner condition was selected as factors in this paper. As known, the outer trigger happens by chance, and they are difficult to be quantitatively as-

---

**Table 3** Result of coefficients (Strong: more sensitive, weak: less sensitive)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Constant</th>
<th>Geology</th>
<th>Clinal-aspect</th>
<th>Slope angle</th>
<th>Land use</th>
<th>Soil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>$a_0$</td>
<td>$a_1$</td>
<td>$a_2$</td>
<td>$a_3$</td>
<td>$a_4$</td>
<td>$a_5$</td>
</tr>
<tr>
<td>Linear</td>
<td>0.009</td>
<td>0.330</td>
<td>0.305</td>
<td>0.238</td>
<td>0.223</td>
<td>0.131</td>
</tr>
</tbody>
</table>

Fig. 5 Statistical Charts of landslide factors

Fig. 6 The model to calculate the coefficients in landslide zones

---

sessed, therefore they were not selected as factors in this paper. So some existing landslide triggered by outer condition would be underestimated. In this study region, there are many and pretty thick snow fall in this area in winter season, and snow melting is easy to occur not only in early spring but also in the mid-winter season (Honda and Nozaki, 1999). Snow melting should be considered as a factor to spatio-temporal patterns of landslides in future for Toyama Prefecture.

Some problems still remain, such as, is the set of coefficients from the Himi District by multiple linear regression applicable to the Toyama region? How does the classification affect the statistical analysis of every factor and then affect the result of landslide hazard assessment?

Acknowledgement

We would like to express many thanks to the president of ARCGEO Inc., Mr. Haruaki Tsushima, his support was important to finish this research work. We
would like to thank Dr. David Morley (Environmental Change Research Center of University College London) who sincerely improved this English paper. We also thank the anonymous reviewers of this paper, their hard work and comments make the quality of this paper improved. The MONBUSHO SCHOLARSHIP funded the study of one of the authors Pan in Japan.

References
Imai, I., Sakamoto, T. and Nozawa, T. (1967a, b) : Geological map of Abugashima (1 : 50,000). Geological Survey of Japan
(Received May 1, 2007. Accepted December 4, 2007)