Available online at www.sciencedirect.com
SCIENCE DIRECT®
Geomorphology 65 (2005) 15–31
GEOMORPHOLOGY
www.elsevier.com/locate/geomorph

The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan

Lulseged Ayalew*, Hiromitsu Yamagishi

Niigata University, Department of Environmental Science, Ikarashi 2-no-cho 8050, Niigata, 950-2181 Japan

Received 16 October 2003; received in revised form 2 June 2004; accepted 7 June 2004
Available online 1 September 2004

Abstract

As a first step forward in regional hazard management, multivariate statistical analysis in the form of logistic regression was used to produce a landslide susceptibility map in the Kakuda-Yahiko Mountains of Central Japan. There are different methods to prepare landslide susceptibility maps. The use of logistic regression in this study stemmed not only from the fact that this approach relaxes the strict assumptions required by other multivariate statistical methods, but also to demonstrate that it can be combined with bivariate statistical analyses (BSA) to simplify the interpretation of the model obtained at the end. In susceptibility mapping, the use of logistic regression is to find the best fitting function to describe the relationship between the presence or absence of landslides (dependent variable) and a set of independent parameters such as slope angle and lithology. Here, an inventory map of 87 landslides was used to produce a dependent variable, which takes a value of 0 for the absence and 1 for the presence of slope failures. Lithology, bed rock-slope relationship, lineaments, slope gradient, aspect, elevation and road network were taken as independent parameters. The effect of each parameter on landslide occurrence was assessed from the corresponding coefficient that appears in the logistic regression function. The interpretations of the coefficients showed that road network plays a major role in determining landslide occurrence and distribution. Among the geomorphological parameters, aspect and slope gradient have a more significant contribution than elevation, although field observations showed that the latter is a good estimator of the approximate location of slope cuts. Using a predicted map of probability, the study area was classified into five categories of landslide susceptibility: extremely low, very low, low, medium and high. The medium and high susceptibility zones make up 8.87% of the total study area and involve mid-altitude slopes in the eastern part of Kakuda Mountain and the central and southern parts of Yahiko Mountain.

Keywords: Landslide; Slope stability; Susceptibility; GIS; Logistic regression; Japan

* Corresponding author. Fax: +81 25 262 7289.
E-mail address: ayalew@env.sc-niigata-u.ac.jp (L. Ayalew).

0169-555X/$ - see front matter
doi:10.1016/j.geomorph.2004.06.010

1. Introduction

Landslide susceptibility mapping relies on a rather complex knowledge of slope movements and their controlling factors. The reliability of landslide susceptibility maps depends mostly on the amount and quality of available data, the working scale and the selection of the appropriate methodology of analysis and modeling. The process of creating these maps involves several qualitative or quantitative approaches (Soeters and van Westen, 1996, Aleotti and Chowdhury, 1999; Guzzetti et al., 1999).

Qualitative methods depend on expert opinions. The most common types of qualitative methods simply use landside inventories to identify sites of similar geological and geomorphological properties

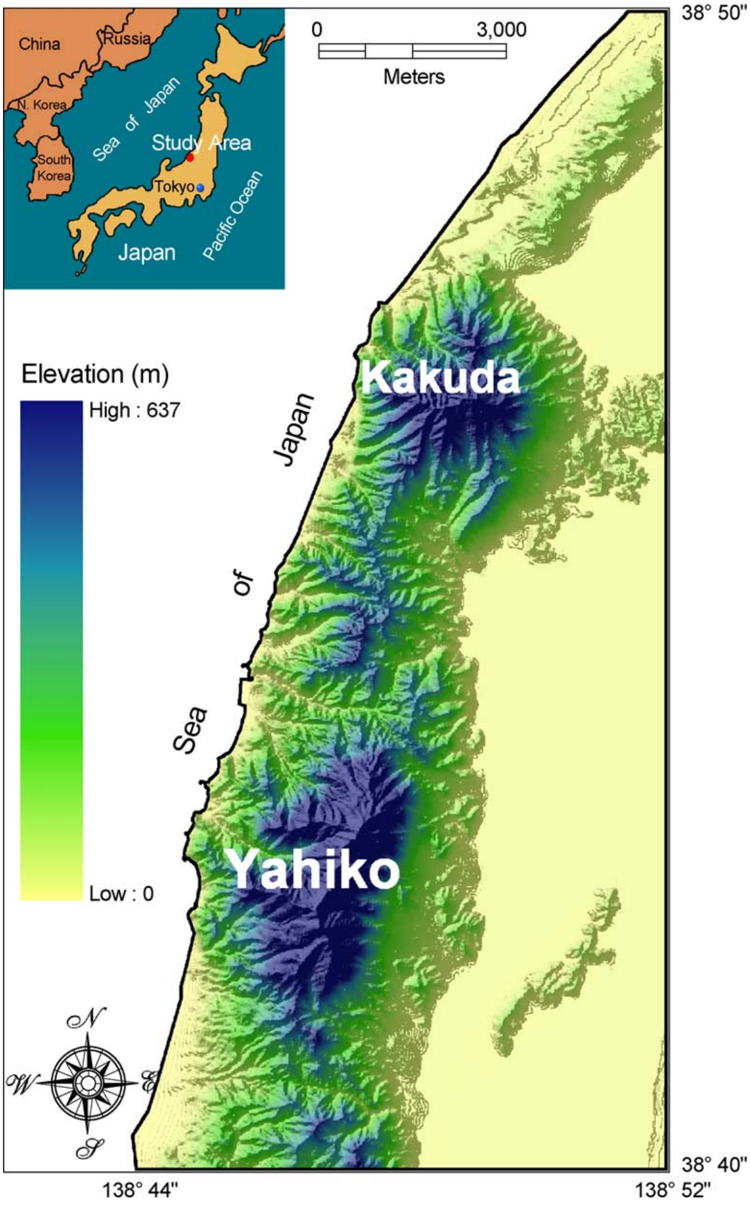

Fig. 1. The study area and its geographic setting.

that are susceptible to failure. Some qualitative approaches, however, incorporate the idea of ranking and weighting, and may evolve to be semiquantitative in nature. Examples are the use of the analytic hierarchy process (AHP) of Saaty (1980) by Barredol et al. (2000) and weighted linear combination (WLC) by Ayalew et al. (2004a,b). AHP involves building a hierarchy of decision elements (factors) and then making comparisons between possible pairs in a matrix to give a weight for each element and also a consistency ratio. It is based on three principles: decomposition, comparative judgment and synthesis of priorities (Malczewski, 1999). WLC is a concept to combine maps of landslide-controlling parameters by applying a standardized score (primary-level weight) to each class of a certain parameter and a factor weight (secondary-level weight) to the parameters themselves. Being partly subjective, results of these approaches vary depending on the knowledge of experts. Hence, qualitative or semiquantitative methods are often useful for regional studies (Soeters and van Westen, 1996; Guzzetti et al., 1999).

Quantitative methods are based on numerical expressions of the relationship between controlling factors and landslides. There are two types of quantitative methods: deterministic and statistical (Aleotti and Chowdhury, 1999). Deterministic quantitative methods depend on engineering principles of slope instability expressed in terms of the factor of safety. Due to the need for exhaustive data from individual slopes, these methods are often effective for mapping only small areas. Landslide susceptibility mapping using either multivariate or bivariate statistical approaches analyzes the historical link between landslide-controlling factors and the distribution of landslides (Guzzetti et al., 1999). Bivariate statistical analyses (BSA) involve the idea of comparing a landslide inventory map with maps of landslide-influencing parameters in order to rank the corresponding classes according to their role in landslide formation. Ranking is normally carried out using landslide densities.

A variety of multivariate statistical approaches (MSA) exist, but those commonly used to map landslide susceptibility include discriminant analyses and logistic regression. Stepwise discriminant analyses have been used by Carrara et al. (1991, 1995, 2003) to classify stable and unstable slope-units in Italy. The method was also reported to be significant to define landslide susceptibility classes in the Spanish Eastern Pyrenees (Baeza and Corominas, 2001). Logistic regression has been applied for susceptibility mapping by various researchers including Bernknopf et al. (1988), Jade and Sarkar (1993), Wieczorek et al. (1996), Atkinson and Massari (1998), Guzzetti et al. (1999), Gorsevski et al. (2000), Lee and Min (2001), Dai et al. (2001), Dai and Lee (2002, 2003) and Ohlmacher and Davis (2003). This study attempts to extend the application of logistic regression because the method requires fewer theoretical assumptions than discriminant analysis. In addition, there is an interest in showing how it can be combined with BSA to simplify the interpretation of the resulting regression function.

2. The study area

The study area encompasses the Kakuda-Yahiko Mountains and their surroundings. It is located about 20 km south of Niigata City in Central Japan (Fig. 1), and is bounded by longitudes of 138°44″E and 138°52″E, and latitudes of 38°40″N and 38°50″N. It covers two adjacent 1:25,000 topographic sheets of the National Geographic Institute of Japan and has an extent of about 105 km^2. Its western and southern limits are respectively dictated by the Sea of Japan and an artificial channel built to divert part of the region's major river (Shinano River) to the sea. The eastern and northern boundaries coincide simply with the limit of the topographic sheets.

The bedrock exposure in the area is dominated by Neogene submarine volcanoclastic rocks and lavas (Nagase et al., 1992). Kakuda Mountain (482 m a.s.l.) is composed of andesitic volcaniclastic rocks and lavas. Around the northern portion of this mountain, andesitic hyaloclastites and their feeder dykes are well exposed (Yamagishi, 2002). Yahiko Mountain (637 m a.s.l.) consists mainly of basaltic and rhyolitic volcanic rocks as well as andesitic hyaloclastites interbedded with mudstones and tuffs (Fujibayashi and Ushiki, 1992). Some andesitic hyaloclastites and massive dolerite sheets are intruded into these rocks. The lowlands surrounding the mountains are generally composed of recent fluvial sediments and colluvium. As a result of the constant supply of moisture from the

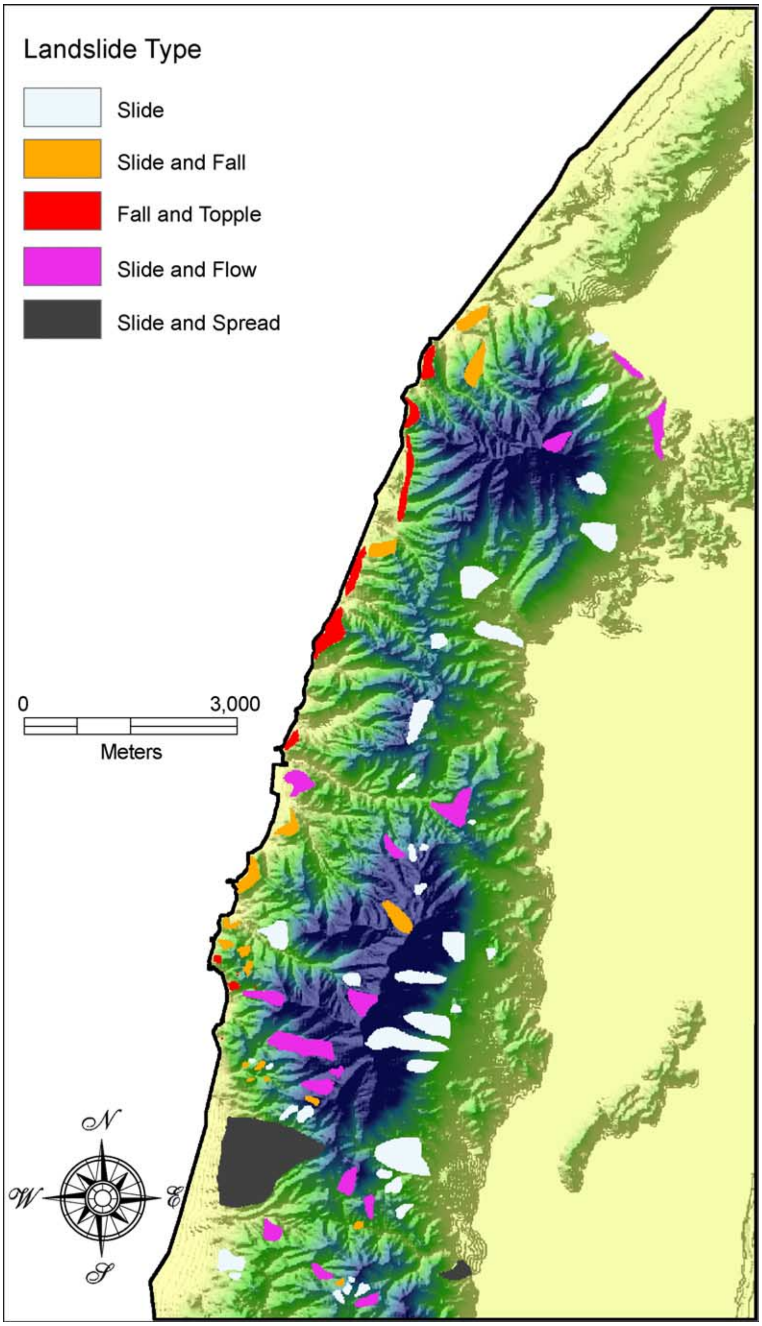

Fig. 2. The landslide inventory map of the study area. Rock falls and topples are dominant along the main highway that runs parallel to the Sea of Japan.

sea and the local fluctuation of groundwater, almost all rocks in the study area have undergone a certain degree of weathering. In many slopes, weathering has penetrated deep into rock masses through joints and bedding planes. The degree of weathering is especially high in rock cliffs and road cuts close to the coast. The rocks in these cliffs and road cuts tend to fret on exposure or breakdown easily to granular constituents when excavated.

The accumulation of volcanic rocks in the Kakuda-Yahiko Mountains built two peaks that can clearly be seen on aerial photographs and digital elevation models. A further flow of a small quantity of lava towards the section between the two peaks formed a NNE–SSW-trending ridge. Due most probably to eruption mechanisms and subsequent geological processes, Kakuda Mountain is more circular than Yahiko Mountain. Accordingly, there is a radial drainage pattern in Kakuda but parallel and subparallel patterns in Yahiko. In the middle, the majority of streams drain towards the west to join the sea although some flow to the north and then east to form a dendritic drainage pattern. Slope recession by minor slumping along streams appears to have affected many of the scarps in the Kakuda-Yahiko Mountains.

Land-use in the area is directed towards locally based social endeavors, and the land's territories are used mainly for settlements and small-scale agro-industrial activities. Vegetation is dense throughout the mountains but scarce in the lowlands. The climate of the study area is humid and temperate. Precipitation is heavy and occasionally intense during typhoons and occurs in the form of snow in winter. The mean yearly precipitation at a station located around 2.3 km south of the study area over the period of 1987–1999 is 1841 mm. Access to the region as a whole is by several interconnected roads.

3. Description of landslides

Landslides were identified from the interpretation of 1:20,000 scale color aerial photographs. A series of field trips were helpful to check the sizes and shapes of landslides, identify the types of movements and the materials involved, and determine the activity (active, dormant, etc.) of failed slopes. In total, 87 landslides were mapped (Fig. 2), and subsequently digitized and rasterized in GIS with a grid size of 10×10 m. This is a dimension dictated by the size of the DEM used to generate geomorphological parameters, namely slope gradient and aspect. Other vector data layers such as lithology, bed rock-slope relationship, lineament and road maps were also rasterized with this size. The mapped landslides cover an area of 5.82 km^2, which is 5.52% of the project site. The smallest landslide that we were able to identify from the aerial photographs and subsequently recognize in the field has an extent of 153 m^2 while the largest one located in the southwestern part of Yahiko Mountain is 1.21 km^2. At the time of mapping, landslide boundaries were limited to detachment zones. But, in some places where there are signs of further movement such as cracks and tilted trees, the upper portions of accumulation zones were also included.

Description and classification of landslides was mainly based on the system of Cruden and Varnes (1996), which takes into consideration the types of movements, the materials involved and the states or activities of failed slopes. In Fig. 2, landslides are portrayed according to the types of movements, namely slide, fall, flow, spread and topple. Field observations revealed that two of these terms need to be combined to appropriately label some landslides. This study separates debris flows from other types of mass movements. Hence, debris-flow deposits present in the eastern part of the Kakuda-Yahiko Mountains are not included in the landslide inventory map. In the study area, the dominant type of failure is "slide" (Fig. 2), and most of the landslides are either suspended or dormant. The materials involved in many landslides are a mixture of soils, gravels and cobbles. Rock falls and topples along the main highway parallel to the Sea of Japan consist of rock blocks of up to 2–3 m in diameter. The regression analysis used in this study is not based on the types of slope failures. Rather, it considers all landslides together in order to minimize the effect of an unequal proportion of 0 and 1 in the dependent variable on the final result, as will be discussed later.

4. Landslide influencing parameters

In logistic regression, the more independent variables are included, the more complete the model will be, but only when they play a major role in determining the dependent variable. Selecting those independent variables with a major role is, however, a

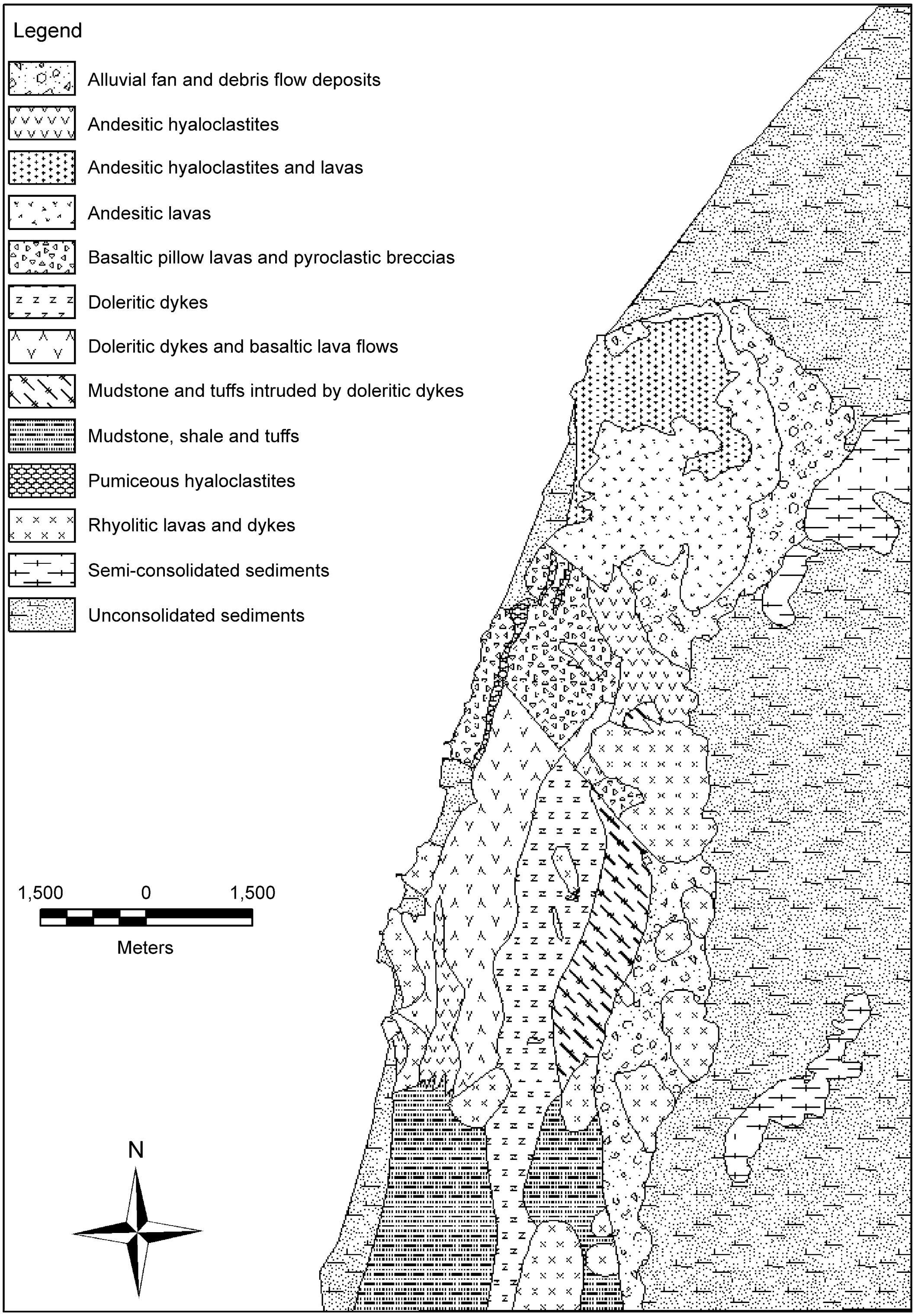

Fig. 3. The lithological map of the study area (modified from Hasegawa, 1973).

Table 1
Landslide densities computed for the parameter lithology

Class	Rock units	Density (%)
1	Alluvial fan and debris flow deposits	7.62
2	Andesitic hyaloclastites	6.67
3	Andesitic hyaloclastites and lavas	7.62
4	Andesitic lavas	2.86
5	Basaltic pillow lavas and pyroclastic breccia	10.48
6	Doleritic dykes	8.57
7	Doleritic dykes and basaltic lava flows	5.71
8	Mudstone and tuffs intruded by doleritic dykes	17.14
9	Mudstone, shale and tuffs	25.71
10	Pumiceous hyaloclastites	0.95
11	Rhyolitic lavas and dykes	5.71
12	Semiconsolidated sediments	0.01
13	Unconsolidated sediments	0.95

difficult task. There are neither universal criteria nor guidelines. The general consensus is that any independent variable must be operational (has a certain degree of affinity with the dependent variable), complete (is fairly represented all over the study area), non-uniform (varies spatially), measurable (can be expressed by any of the different types of measuring scales) and non-redundant (its effect should not account for double consequences in the final result).

In the study area, the main triggering factor for landsliding is the high amount of precipitation. However, the regression analysis does not include precipitation because rain and snowfall are relatively uniform throughout the study area. The same can be said of seismicity. The contrast in the density of vegetation within the mountains where landslides occur is also hardly visible. Hence, seven landslide-influencing parameters were selected: lithology, bed rock-slope relationship, lineaments, elevation, slope gradient, aspect and proximity to roads.

4.1. Geological parameters

It is widely recognized that geology greatly influences the occurrence of landslides, because lithological and structural variations often lead to a difference in strength and permeability of rocks and soils. In the Kakuda-Yahiko Mountains and their surroundings, field surveys indicated that the lithology, bed rock-slope relationships and lineaments determine the boundaries of landslides and their overall distribution. The Kakuda-Uchino geological map, produced for the study area and its surroundings by Hasegawa (1973), consists of more than 25 rock units. Based on lithological similarities, these rock units were grouped into 13 categories (Fig. 3). In GIS, the rasterized form of this lithological map was compared with the landslide inventory map to calculate landslide densities on the basis of BSA. The calculation process first takes the ratio between the area occupied by landslide pixels on a class of a certain parameter and the total area of that class. This was repeated for up to the number of classes available in that parameter. Then, these ratios were added and each ratio was divided by the total sum to obtain landslide density.

Table 1 presents the landslide densities computed for the 13 classes of the parameter "lithology". The class representing "mudstone, shale and tuff" has the highest landslide density (25.71%). The rock units of this class occur along the mid-elevation of Yahiko Mountain. On the other hand, the unconsolidated and semiconsolidated deposits in the lowlands have a negligible amount of landslide densities.

Lineaments were extracted from a mosaic of 1:20,000 scale aerial photographs using edge enhancement and filtering techniques as well as subsequent field verifications. Lineaments may be identified manually using an interactive system or automatically with the help of pattern recognition (Chorowicz et al., 1989), edge detection routines (Heddi et al., 1999) or texture analyses (Morris, 1991). The study of lineaments may help revealing generalities that may assist in understanding the cause of landslides in the region. Lineaments include tectonic structures and geomorphologic signatures such as topographic breaks. In spite of dense vegetation in the mountains, 103 lineaments were extracted in this study aided by abrupt relief changes, valleys and cliffs with sharp tonal contrasts, lithological variations, erosional features and changes in drainage patterns. For the purpose of a landslide susceptibility analyses, the lineaments were enclosed by 100 m buffer zones. This diameter is an average threshold set based on a comprehensive assessment of how far slope failures extend from mountain scarps, topographic breaks and any other linear futures. With an assumption of having landslides along lineaments, landslide density was assumed to be 100% within the buffer zones and 0% outside.

Table 2
The bed rock-slope relationships and their corresponding landslide densities

Class	Bed rock-slope relationship	Density (%)
1	Flat	0.01
2	Under-dip slope	35.48
3	Dip slope	29.03
4	Over-dip slope	35.48

Geometric relationship between rock beds and slope can influence the condition of groundwater flow, and hence, the process of landsliding (Cruden, 1989; Cruden and Hu, 1996). In general, if rocks are dipping in the same direction as the topographic surface, the slope is said to be cataclinal. If the beds dip in the direction opposite to the latter, anaclinal-slopes are created. When the strike of rock beds is perpendicular to the azimuth of mountain faces, the outcome are orthoclinal-slopes. In the study area, the terrain has established three types of cataclinal-slopes (over-dip, dip and under-dip) and a flatland. At higher grounds, the topographic surfaces are steeper than bedding planes and over-dip slopes are formed. In low altitudes, it is the opposite and under-dip slopes are common. In the middle, there are zones for dip-slopes where the slopes and rock beds are inclined parallel to each other. High possibility of failure exists in over-dip slopes where bedding planes daylight on the topographic surface, although landslide density is identical in the study area for both over-dip and under-dip slopes as shown in Table 2. The likelihood of landslide occurrence decreases in dip-slopes and becomes negligible in flatlands.

Table 3
Elevation classes and landslide densities

Class	Elevation (m)	Density (%)
1	0–50	1.68
2	51–100	9.18
3	101–150	10.94
4	151–200	15.38
5	201–250	11.27
6	251–300	9.73
7	301–350	8.91
8	351–400	7.64
9	401–450	7.36
10	451–500	7.04
11	501–550	6.77
12	551–600	3.53
13	601–637	0.57

4.2. Geomorphological parameters

Surface topography, by controlling flow sources, flow direction and soil moisture concentration, is an important factor that limits the density and spatial extent of landslides. Therefore, significant terrain attributes such as slope gradient and aspect were derived from a 10×10 m DEM. The use of DEMs as a basis for deriving secondary parameters used for geomorphological analyses has been well documented by Moore et al. (1991). GIS technology permits patterns of slope instability to be analyzed at the scale of the DEM.

Elevation is useful to classify the local relief and locate points of maximum and minimum heights within terrains. To calculate landslide densities for different relief classes, the relief map was divided into 13 altitude classes on 50-m basis. The highest density corresponds to the class with an elevation range of 150–200 m (Table 3), where many slope cuts are found in the mountains. Density decreases both upward and downward from this elevation range.

Slope gradient is the most substantial cause of landsliding. At local scales, it affects the concentration of moisture and the level of pore pressure, and is often useful to resolve detailed patterns of instability. At larger scales, it controls regional hydraulic continuity, and is considered as an important factor for GIS-based landslide susceptibility mapping by Guzzetti et al. (1999), Lee and Min (2001), Dai and Lee (2002) and Ohlmacher and Davis (2003). Here, we defined five categories of slope angles and computed the corresponding landslide densities. Flat areas with a slope angle of less than 3° harbor no landslides and have a negligible landslide density (Table 4). Landslide density is high (34.22%) for slopes with a gradient of 4–15°.

Aspect reveals the influence that strong waves, differential weathering and subsequent coastal erosion

Table 4
Five classes of slope angle and landslide densities

Class	Slope (°)	Density (%)
1	0–3	2.56
2	4–15	34.22
3	16–30	33.03
4	31–45	19.98
5	46–68	10.21

Table 5
Landslide densities obtained for aspect

Class	Aspect	Density (%)
1	Flat	0.00
2	North	12.00
3	Northeast	14.00
4	East	10.00
5	Southeast	7.00
6	South	7.00
7	Southwest	12.00
8	West	22.00
9	Northwest	16.00

have on landslide distributions in localities close to oceans. In the study area, west-facing slopes near the coast of the Sea of Japan are affected by many landslides. Computed landslide densities agree with this observation and slopes inclined to the west have a landslide density of 22% (Table 5). In contrast, slopes with east, south and southeast orientations possess low landslide densities since they meet little wet and dry cycles, conditions that discourage landslide formations.

4.3. Proximity to roads

Road-cuts are usually sites of anthropologically induced instability. A given road segment may act as a barrier, a net source, a net sink or a corridor for water flow, and depending on its location in the mountains, it usually serves as a source of landslides. Some slope failures start above roads but are often intercepted by them. For this reason, roads are included in GIS-based landslide susceptibility analyses (e.g. Larsen and Parks, 1997; Shaban et al., 2001). The complete transportation network in the study area is about 170 km in length. Excluding roads on flatlands and on top of ridges, we selected approximately 47 km of roads to be considered in this study. This includes tunnels, seaside highways and roads in the mountains.

The traditional method in most GIS-based studies for considering the effect of roads on landslides is to construct buffers around them (Larsen and Parks, 1997). While the same approach was adopted in this study, buffers were built not according to the "distance to the center of roads" methodology. In the Kakuda-Yahiko Mountains, cliffs are present only on one side of the roads. Along the highway that runs parallel to the Sea of Japan for example, the western side is a slightly inclined surface that falls sharply to the sea. Only the eastern side is walled up by long cliffs of Kakuda Mountain. In the GIS analysis, this demands the creation of "half buffers" that stretch only on one side of the roads.

Before creating buffers, however, there was a need to determine buffer length thresholds. Field observations demonstrated that landslide frequency is greatest within about 40 m of roads in mountains and is relatively high as far as 100 m in the case of coastal highways, and 150 m in sectors where tunnels are present. Therefore, "half-buffers" with an interval of 50 m on the side of roads in mountains, 100 m along coastal highways and 150 m around tunnels were constructed. As in the case of lineaments, landslides density was assumed to be 100% within buffered roads.

5. The application of logistic regression

The principle of logistic regression (LR) rests on the analysis of a problem, in which a result measured with dichotomous variables such as 0 and 1 or true and false, is determined from one or more independent factors (Menard, 1995). In the case of landslide susceptibility mapping, the goal of LR would be to find the best fitting (yet reasonable) model to describe the relationship between the presence or absence of landslides (dependent variable) and a set of independent parameters such as slope angle, aspect and lithology. LR generates the model statistics and coefficients of a formula useful to predict a logit transformation of the probability that the dependent variable is 1 (probability of occurrence of a landslide event). It does not define susceptibility directly but an inference can be made using the probability. Generally, LR involves fitting the dependent variable using an equation of the form:

$$Y = \mathrm{Logit}(p) = \ln(p/(1-p))$$
$$= C_0 + C_1X_1 + C_2X_2 + \ldots + C_nX_n \quad (1)$$

Where p is the probability that the dependent variable (Y) is 1, $p/(1-p)$ is the so-called odds or likelihood ratio, C_0 is the intercept, and C_1, C_2, …, C_n are coefficients, which measure the contribution of independent factors ($X_1, X_2, \ldots, X_n$) to the variations in Y.

In order to appropriately interpret the meaning of Eq. (1), one has to use the coefficients as a power to the natural log(e). The result represents the odds ratio or the probability that an event will occur divided by the probability that it fails to do so. If a coefficient is positive, its transformed log value will be greater than one, meaning that the event is more likely to occur. If a coefficient is negative, the latter will be less than one and the odds of the event occurring decreases. A coefficient of 0 has a transformed log value of 1, and it does not change the odds one way or the other. For a positive coefficient, the probability plotted against the values of an independent variable follows an S-shaped curve. A mirror image will be obtained for a negative coefficient (Menard, 1995).

5.1. Analytical approaches

In LR or even in linear regression, it does little good to combine data with different measuring scales. So, the first step ahead of the main statistical analyses is to make sure that data have been normalized in a manner LR needs. Failure to do so generally leads to problems during the interpretation of the final results. In the application for landslide susceptibility mapping, the common solution is to create layers of binary values (dummy variables) for each class of an independent parameter (Guzzetti et al., 1999; Lee and Min, 2001; Dai et al., 2001; Dai and Lee, 2002; Ohlmacher and Davis, 2003). If the number of parameters is small, this approach is good. If there are many parameters, however, it would produce a long regression equation and may even create numerical problems. It may also work against some basic assumptions of LR such as the absence of strong correlations among independent variables (multicollinearity). It might also be difficult to understand statistical results and evaluate the role of each independent variable in the final model.

In this study, we used a different approach. We first arranged classes of all parameters according to their corresponding landslide densities computed earlier using bivariate statistical analyses. The use of landslide densities allowed us to express the independent parameters by the same scale. Then, a regression was performed among the independent parameters. At the end, a class with a high landslide density, which corresponds to a parameter having a higher positive coefficient, was considered to play a greater role in causing landslides. In the parameter “lithology”, for example, landslide density is high for the class containing mudstone, shale and tuffs, and low for the semiconsolidated deposits. If this parameter has a coefficient greater than 0 after regression, then the probability for a landslide to occur in an area where there are mudstone, shale and tuffs is higher than in a place where semiconsolidated deposits are exposed. In general, the approach used here reduces the length of the regression equation and avoids multicollinearity.

There are also differences in the literature in how coefficients are computed and assigned among different classes of a certain parameter. For example, Lee and Min (2001) assigned a coefficient of 0 to the last class of each parameter. This means, if granite was the last class in the parameter “lithology”, then it would automatically get a coefficient of 0. Dai and Lee (2002) “overrode” the coefficients corresponding to these last classes because they are used as default reference categories. The approach of Ohlmacher and Davis (2003) was such that coefficients belonging to the classes of the parameter “geological units” were constrained to sum to 0 so that the last class, shale in their case, could get what was remained from the addition and subtraction processes. This problem was not an issue in this study because, as mentioned above, regression was conducted among parameters, not classes of parameters.

Another issue where there are major differences in literature is the size of samples taken to create the dependent variable. It is generally recommended in LR to use equal proportions of 1 (“landslide”) and 0 (“no-landslide”) pixels, but this was not usually the case in many works. For example, Atkinson and Massari (1998) used training data coming from 1.2% of the area under investigation and unequal proportions of 1 and 0 pixels. Dai and Lee (2002) adopted the same approach, but used equal numbers of pixels. Ohlmacher and Davis (2003) considered data from their entire project site, and hence, unequal pixel proportions. This might also be the case in the work of Guzzetti et al. (1999). Here, we used data from all over the area because landslide densities were determined based on the entire study domain. This

Table 6
Summary statistics of the logistic regression model

Statistics	Value
Total number of pixels	1,054,768
$-2\ln L$ (L=likelihood)	447,111.2188
$-2\ln L_0$	359,804.6250
Model chi-square	87,306.5938
Goodness of fit	770,570.7500
Pseudo R^2	0.1953
ROC	0.8358

undoubtedly leads to unequal pixel proportions, since a considerable portion of the project site is flat and takes 0 in the dependent variable. To minimize the effect of the unequal proportion, we considered all landslides together and increased the proportion of 1.

5.2. Statistical results and discussions

Table 6 summarizes the overall model statistics of the regression conducted in this study using IDRISI. A key starting point could be the model chi-square whose value provides the usual significance test for LR. It is a difference between $-2\ln L$ (L=likelihood) for the best-fitting model and $-2\ln L_0$ for the null hypothesis in which all the coefficients are set to 0, and measures the improvement in fit that the independent variables brought into the regression. In our case, the high value for model chi-square indicates that the occurrence of landslides is far less likely under the null hypothesis (without landslide influencing parameters) than the full regression model (where the parameters are included). The goodness of fit is an alternative to model chi-square for assessing the significance of LR models. It is calculated based on the difference between the observed and the predicted values of the dependent variable. The smaller is this statistic, the better fit it indicates.

The pseudo R^2 value, which can be calculated from $1-(\ln L/\ln L_0)$, cautiously indicates how the logit model fits the dataset (Menard, 1995). Thus, pseudo R^2 equal to 1 indicates a perfect fit, whereas 0 shows no relationship. When a pseudo R^2 is greater than 0.2, it shows a relatively good fit (Clark and Hosking, 1986). The pseudo R^2 in this study is closer to 0.2. An alternative approach, which is much easier to interpret, is to look at how well the model actually predicts the dependent variable. In this case, IDRISI uses the so-called Relative Operating Characteristic (ROC) to compare a Boolean map of "reality" (the presence or absence of landslides) with the probability map. The ROC value ranges from 0.5 to 1, where 1 indicates a perfect fit and 0.5 represents a random fit. A value of 0.8358 is obtained in this study, which can be taken as a sign of good correlation between the independent and dependent variables.

As it was discussed above, the relative importance of the independent variables can be assessed using the corresponding coefficients in the LR model. In this study, all coefficients except the one belonging to lineaments are positive (Table 7), indicating that they are positively related to the probability of landslide formation through the log transformation. Specially, the coefficient that belongs to the parameter "proximity to roads" strongly departs from 0 and led to the inference that the road network has a higher effect on the development of landslides than any other parameter. The parameter "rock bed-slope relationship" is also an important factor to reckon with. Among the geomorphological parameters, "aspect" plays a good role followed by "slope angle". However, the regression model differs from our observation that elevation is a good estimator of slope cuts that can be places of rock falls and topples. In fact, its coefficient is close to 0, which indicates that it has a little impact on the occurrence of landslides. Actually, the role of elevation has already been emphasized in the model by other parameters such as road cuts and rock bed-slope relationship with which it has a strong correlation, and the model has simply tried to avoid double consequences.

In general, mountainous roads open previously less accessible areas to greater economic development, but may also pose serious slope stability problems in

Table 7
The regression coefficients obtained for the seven independent parameters

Independent parameter	Coefficient
Aspect	0.22384946
Elevation	0.09522519
Bed rock-slope relationship	0.38681456
Lineaments	−0.51347333
Lithology	0.20558186
Slope angle	0.21687476
Proximity to roads	0.74134341
Intercept	−8.03402433

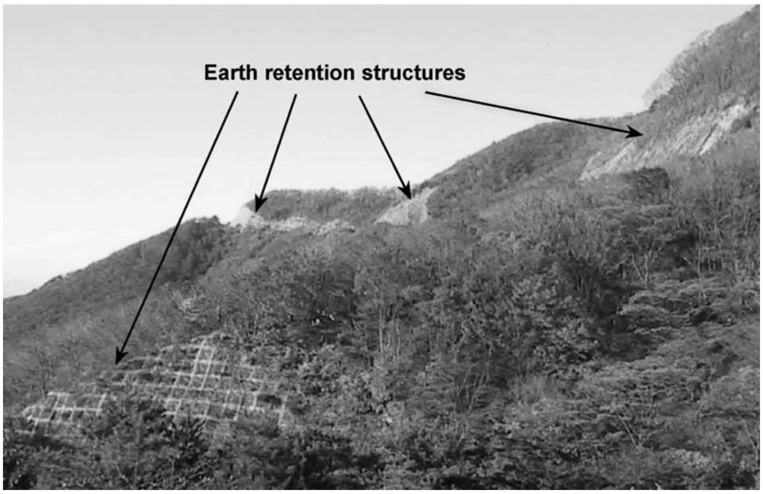

Fig. 4. Reinforced road-cuts in the southwestern part of Yahiko Mountain.

many places although this is conditioned on their location, construction practices, geology, geomorphology and climate. This has been reported by many researchers (Anderson, 1983; Haigh et al., 1993; Larsen and Parks, 1997; Sah and Mazari, 1998; Shaban et al., 2001; Ayalew and Yamagishi, 2004). Chang and Slaymaker (2002) especially reported how a landslide density has increased in Western Foothills of Taiwan after a construction of a major highway. In the Kakuda-Yahiko Mountains, many of the landslides consist of sections of roads or occurred close to them. In fact, most of the rock falls and rock topples in the region occurred along regional highways. In addition, dynamic processes associated with road networks brought obvious changes to the landscape in the last several decades. This has for a long time demanded an application of a dense retaining structure of the type shown in Fig. 4. Therefore, the LR model discussed above captured what has been observed in the field and tells how important road cuts are in timing the occurrence of landslides. The model is not a result of bias originated from the desire to map more accurately cases along roads. It must, therefore, be taken as important information, if providing long-term stability to the mountains is being sought.

5.3. Predicted probabilities and the susceptibility map

In addition to the model statistics and coefficients, the final result of the regression process in IDRISI is a predicted map of probability defined by numbers that are constrained to fall between 0 and 1. The more these numbers are close to 1, the better they indicate the likelihood of finding the mapped landslides. A threshold of 0.5 can also be used to distinguish the two cases where landslide has and has not occurred (Dai and Lee, 2002). Fig. 5 presents the probability map obtained in this study. Most of the areas have a pixel value of 0, reflecting the unequal proportion of samples taken for the dependent variable. But, the number of pixels above 0.5 is also significant and the attempt for correctly classifying the presence and absence of landslides is relatively good.

In seeking a susceptibility map, the method adopted in literature is to divide the histogram of the probability map into different categories based on expert opinions (Guzzetti et al., 1999; Lee and Min, 2001; Dai and Lee, 2002; Ohlmacher and Davis, 2003). This type of changing continuous data into two

Table 8
The classification system used to produce landslide susceptibility categories

Probability range	Class name	Coverage (%)
0–0.02	Extremely low susceptible	46.11
0.03–0.09	Very low susceptible	33.66
0.1–0.16	Low susceptible	11.36
0.17–0.23	Medium susceptible	4.73
0.24–0.89	High susceptible	4.14

or more categories does not take into account the relative position of a case within the probability map and is neither fully automated nor statistically tested. In this study, we considered four classification systems that use quantiles, natural breaks, equal intervals and standard deviations, and attempted to choose the one that best suits the information and the scale of our investigation.

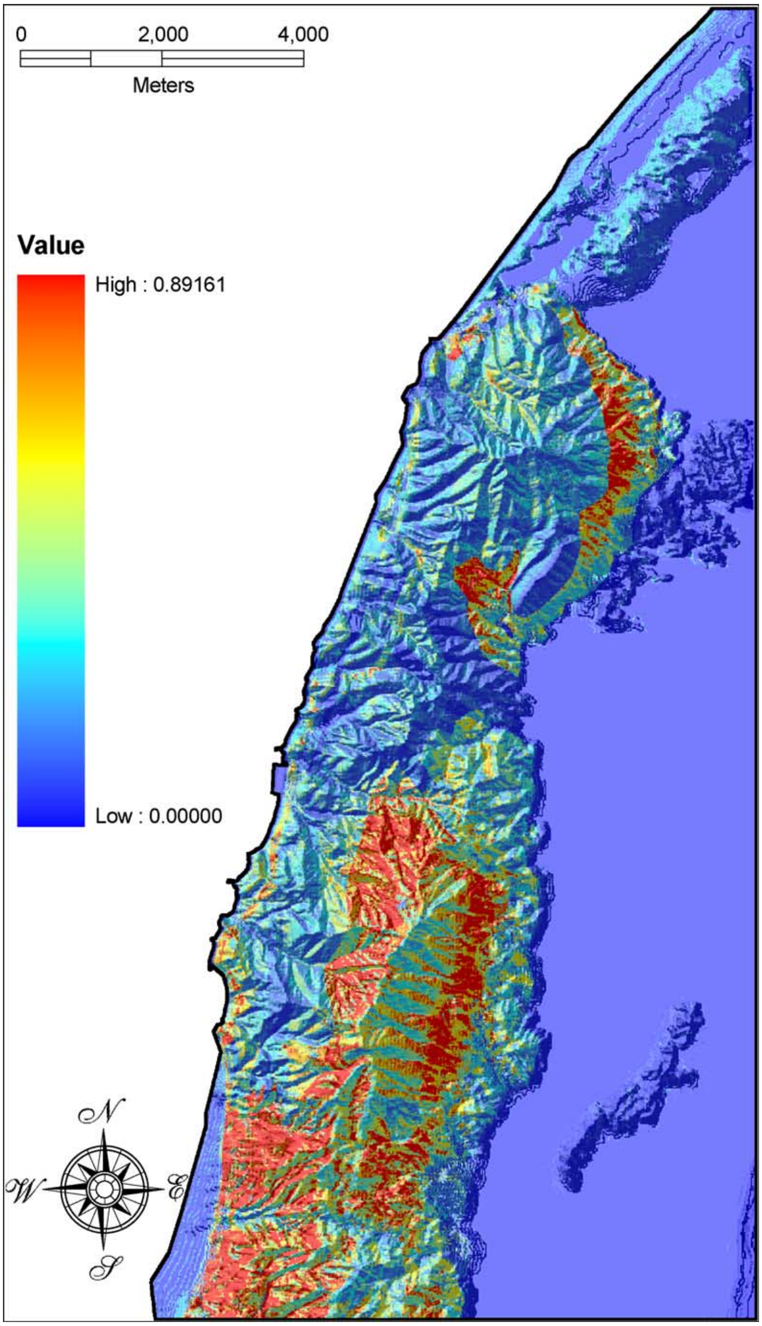

Fig. 5. The probability map obtained after the logistic regression analyses.

A few trials showed that quantile-based classification system has a disadvantage in that it places widely different values into the same class. The use of natural breaks is good when there are big jumps in data values, which is not the case in our probability map. Using equal intervals was also found to be not helpful because it emphasizes one class relative to others. The standard deviation method has a certain

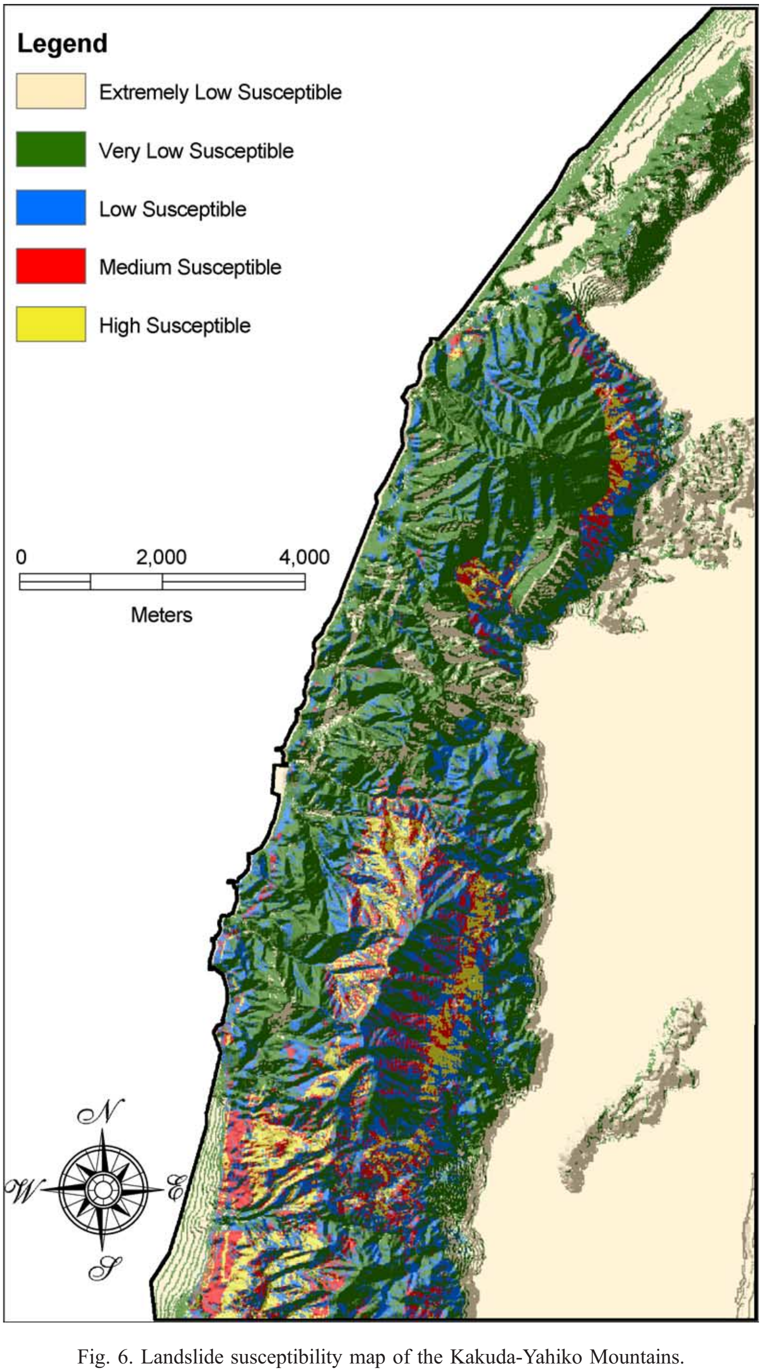

Fig. 6. Landslide susceptibility map of the Kakuda-Yahiko Mountains.

merit in that it uses the mean to generate class breaks, and allowed us to divide the probability map into five categories of landslide susceptibility: extremely low, very low, low, medium and high (Table 8), by adding or subtracting one standard deviation at a time. As shown in Fig. 6, 46.11% of the study area is designated to be extremely low susceptible. Very low, low and medium susceptible zones make up 33.66%, 11.36% and 4.73%, respectively. The zone corresponding to high susceptibility constitutes 4.14%, and its near similarity with the total coverage of the mapped landslides (5.52%), shows the validity of the system adopted to divide the probability map.

According to the above classification, most parts of Kakuda Mountain are found in extremely low, very low or low susceptible zones. A few land parcels in the eastern part are susceptible at medium scale, but high susceptible zones are rare. In the center of the study area, the proportion of the high susceptible zones increased significantly (Fig. 6). In Yahiko and further south, susceptibility is high in the middle sections of the east and west sides of the mountain. In the western flank that faces the Sea of Japan especially, mid-altitude and some low-elevated areas, covered by relatively weak rocks such as mudstone and shale, are classified as high or medium susceptible to the process of landsliding.

In general, depending on the independent parameters considered, the landslide inventory map and the statistical approach used, the best predictor parameters and the predicted probability map of a logistic regression can vary considerably. So far, there is no standard which limits the number of independent parameters, and actually, there is no need to have one either, as influencing factors can vary on the basis of the study area characteristics. Hence, elevation and slope angle were the best predictor variables for estimating the probability of landslide occurrences in the study of Dai and Lee (2002) and lithology in Ohlmacher and Davis (2003). The model in this study showed that the "proximity to roads" parameter has a strong positive correlation with the predicted map of probability. These variations in best predictors indicate the trade-offs that logistic regression makes between input and output, and are actually good signs of the merit that the method has in satisfying specific observations.

6. Conclusion

The preparation of landslide susceptibility map is a major step forward in hazard management. Nowadays, such maps can be prepared by GIS-based qualitative and quantitative techniques. There are many qualitative and quantitative techniques useful to analyze the relationship between landslides and their influencing parameters. This study used logistic regression to prepare a susceptibility map for the Kakuda-Yahiko Mountains, Central Japan. At first, a total of 87 landslides were mapped using aerial photographs and subsequent field checks. The influencing parameters considered include lithology, bed rock-slope relationship, lineaments, slope angle, aspect, elevation and proximity to roads. Later, landslide densities were determined using bivariate statistical analyses.

The first results of the logistic regression were the model statistics and coefficients, which were useful to asses the accuracy of the regression function and the role of parameters on the presence or absence of landslides. The "proximity to roads" parameter was found to have the strongest relationship with slope failures, whereas lineaments portrayed a negative correlation and elevation showed a little role in timing landslide occurrences. The probability map, which was the second outcome of the regression process, was found to have more 0 values in a 0–1 range. But, the map also showed a significant number of pixels above a threshold of 0.5. Using this map, an inference was done to divide the study area into five zones of landslide susceptibility, namely extremely low, very low, low, medium and high. Although the proportion of medium and high susceptible zones was far smaller than the very low and extremely low counterparts, the resulting map supported the observation that landslides are more common in Yahiko than Kakuda and in mid-altitude slopes than in highlands and lowlands.

Acknowledgments

The study was financially supported by the Japanese Society for Promotion of Science (JSPS) and Niigata University. Prof. Robin Chowdhury, Prof. Mauro Soldati, one anonymous reviewer and the editor Prof. Takashi Oguchi provided detailed and helpful comments, for which we are grateful.

References

Aleotti, P., Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. Bulletin of Engineering Geology and the Environment 58, 21–44.

Anderson, M.G., 1983. Road-cut slope topography and stability relationships in St. Lucia, West Indies. Applied Geography 3, 104–114.

Atkinson, P.M., Massari, R., 1998. Generalized linear modeling of susceptibility to landsliding in the central Apennines, Italy. Computers and Geosciences 24, 373–385.

Ayalew, L., Yamagishi, H., 2004. Slope movements in the Blue Nile basin, as seen from landscape evolution perspective. Geomorphology 57, 95–116.

Ayalew, L., Yamagishi, H., Ugawa, N., 2004. Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. Landslides 1, 73–81.

Ayalew, L., Yamagishi, H., Watanabe, N., Marui, H., 2004. Landslide susceptibility mapping using a semi-quantitative approach, a case study from Kakuda-Yahiko Mountains, Niigata, Japan. In: Free, M., Aydin, A. (Eds.), Proceedings of the 4th Asian Symposium on Engineering Geology and the Environment, Geological Society of Hong Kong, vol. 7, pp. 99–105.

Baeza, C., Corominas, J., 2001. Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. Earth Surface Processes and Landforms 26, 1251–1263.

Barredol, J.I., Benavidesz, A., Herhl, J., van Westen, C.J., 2000. Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin, Gran Canaria Island, Spain. International Journal of Applied Earth Observation and Geoinformation 2, 9–23.

Bernknopf, R.L., Cambell, R.H., Brookshire, D.S., Shapiro, C.D., 1988. A probabilistic approach to landslide hazard mapping in Cincinnati, Ohio, with applications for economic evaluation. Bulletin of the International Association of Engineering Geology 25, 39–56.

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., Reichenbach, P., 1991. GIS techniques and statistical models in evaluating landslide hazard. Earth Surface Processes and Landforms 16, 427–445.

Carrara, A., Cardinali, M., Guzzetti, F., Reichenbach, P., 1995. GIS technology in mapping landslide hazard. In: Carrara, A., Guzzetti, F. (Eds.), Geographical Information Systems in Assessing Natural Hazards. Kluwer Academic Publisher, Dordrecht, The Netherlands, pp. 135–175.

Carrara, A., Crosta, G., Frattini, P., 2003. Geomorphological and historical data in assessing landslide hazard. Earth Surface Processes and Landforms 28, 1125–1142.

Chang, J., Slaymaker, O., 2002. Frequency and spatial distribution of landslides in a mountainous drainage basin: Western Foothills, Taiwan. Catena 46, 285–307.

Chorowicz, J., Kim, J., Manoussis, S., Rudant, J.P., Foin, P., Veillet, I., 1989. A new technique for recognition of geological and geomorphological patterns in digital terrain models. Remote Sensing of the Environment 29, 229–239.

Clark, W.A., Hosking, P.L., 1986. Statistical Methods for Geographers. John Wiley and Sons, New York. 518 pp.

Cruden, D.M., 1989. Limits to common toppling. Canadian Geotechnology Journal 27, 737–742.

Cruden, D.M., Hu, X.Q., 1996. Hazardous modes of rock slope movement in the Canadian Rockies. Environmental and Engineering Geoscience 2, 507–516.

Cruden, D.M., Varnes, D.J., 1996. Landslide types and processes. In: Turner, K.A., Schuster, R.L. (Eds.), Landslides: investigation and mitigation, Transport Research Board Special Report, vol. 247, pp. 36–75.

Dai, F.C., Lee, C.F., 2002. Landslide characteristics and slope instability modeling using GIS Lantau Island, Hong Kong. Geomorphology 42, 213–238.

Dai, F.C., Lee, C.F., 2003. A spatiotemporal probabilistic modeling of storm-induced shallow landsliding using aerial photographs and logistic regression. Earth Surface Processes and Landforms 28, 527–545.

Dai, F.C., Lee, C.F., Li, J., Xu, Z.W., 2001. Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. Environmental Geology 40, 381–391.

Fujibayashi, N., Ushiki, H., 1992. Clastic veins filling the cooling joints of dolerite sheet. Contribution of the Department of Geology, Niigata University 7, 73–84.

Gorsevski, P.V., Gessler, P., Foltz, R.B., 2000. Spatial prediction of landslide hazard using logistic regression and GIS. 4th Int. Conference on Integrating GIS and Environmental Modeling, Alberta, Canada. 9 pp.

Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P., 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, central Italy. Geomorphology 31, 181–216.

Haigh, M.J., Rawat, J.S., Bartarya, S.K., Rawat, M.S., 1993. Factors affecting landslide morphology along new highways in the Central Himalaya, Transactions. Japanese Geomorphological Union 14, 99–123.

Hasegawa, Y., 1973. Surficial geological map of the Kakuda-Uchino area (1:50000) with explanatory text. Niigata Prefecture, Niigata, Japan. 95 pp.

Heddi, M., Eastaff, D.J., Petch, J., 1999. Relationships between tectonic and geomorphological linear features in the Guadix-Baza Basin, Southern Spain. Earth Surface Processes and Landforms 24, 931–942.

Jade, S., Sarkar, S., 1993. Statistical models for slope stability classification. Engineering Geology 36, 91–98.

Larsen, M.C., Parks, J.E., 1997. How wide is a road? The association of roads and mass movements in a forested montane environment. Earth Surface Processes and Landforms 22, 835–848.

Lee, S., Min, K., 2001. Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental Geology 40, 1095–1113.

Malczewski, J., 1999. GIS and Multi-Criteria Decision Analysis. John Willey and Sons, New York. 392 pp.

Menard, S., 1995. Applied logistic regression analysis. Sage University Paper Series on Quantitative Applications in Social Sciences, vol. 106. Thousand Oaks, California. 98 pp.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modeling: a review of hydrological, geomorphological and biological applications. Hydrological Processes 5, 3–30.

Morris, K., 1991. Using knowledge-base rules to map three-dimensional nature of geological features. Photogrammetric Engineering and Remote Sensing 57, 1209–1216.

Nagase, K., Kurokawa, K., Honma, H., Abe, Y., Sugino, S., 1992. Neogene volcano-stratigraphy and volcanic history of the Kakudayama-Maze area, Niigata Prefecture. Contribution of the Department of Geology Mining, Niigata University 7, 41–55.

Ohlmacher, C.G., Davis, C.J., 2003. Using multiple regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Engineering Geology 69, 331–343.

Saaty, T.L., 1980. The Analytical Hierarchy Process. McGraw Hill, New York. 350 pp.

Sah, M.P., Mazari, R.K., 1998. Anthropogenically accelerated mass movement, Kulu Valley, Himachal Pradesh, India. Geomorphology 26, 123–138.

Shaban, A., Khawlie, M., Bou Kheir, R., Abdallah, C., 2001. Assessment of road instability along a typical mountainous road using GIS and aerial photos, Lebanon, Eastern Mediterranean. Bulletin of Engineering Geology and the Environment 60, 93–101.

Soeters, R., van Westen, C.J., 1996. Slope instability recognition, analysis, and zonation. In: Turner, K.A., Schuster, R.L. (Eds.), Landslides: investigation and mitigation, Transport Research Board Special Report, vol. 247, pp. 129–177.

Wieczorek, G.F., Gori, P.L., Jager, S., Kappel, W.M., Negussey, D., 1996. Assessment and management of landslide hazards near Tully Valley landslide, Syracuse, New York, USA. Proc. 7th Int. Symposium on Landslides, Trondheim. Balkema, Rotterdam, pp. 411–416.

Yamagishi, H., 2002. Occurrences and the criteria for recognition of the sub-aqueous volcanic rocks on the Kakuda-Yahiko Beaches, Niigata Prefecture. Journal of the Japanese Association for Petroleum Technology 67, 594–600.