# Landslide Hazard Mapping and its Evaluation Using GIS: An Investigation of Sampling Schemes for a Grid-Cell Based Quantitative Method

Amod Sagar Dhakal, Takaaki Amada, and Masamu Aniya

## Abstract

*An application of GIS for landslide hazard assessment using multivariate statistical analysis, mapping, and the evaluation of the hazard maps is presented. The study area is the Kulekhani watershed (124 km²) located in the central region of Nepal. A distribution map of landslides was produced from aerial photo interpretation and field checking. To determine the factors and classes influencing landsliding, layers of topographic factors derived from a digital elevation model, geology, and land use/land cover were analyzed by quantification scaling type II (discriminant) analysis, and the results were used for hazard mapping. The effects of different samples of landslide and non-landslide groups on the critical factors and classes and subsequently on hazard maps were evaluated. Simple random sampling was used to obtain samples of the landslide group, and either an unaligned stratified random sampling or an aligned systematic sampling method generated the non-landslide group. For the analysis, one set of the landslide group was combined with each of five different sets of the non-landslide groups. Combinations of different samples yielded some minor differences in the critical factors and classes. The geology was found to be the most important factor for landslide hazard. The scores of the classes of the factors quantified by the five analyses were used for the hazard mapping in the GIS, with four levels of relative hazard classes: high, moderate, less, and least. The evaluation of five hazard maps indicated higher accuracy for the combinations in which the non-landslide group was generated by the unaligned stratified random sampling method. The agreements in the hazard maps, produced from different sample combinations using unaligned stratified random sampling for selecting non-landslide group, were within the acceptable range for the practical use of a hazard map.*

## Introduction

### Background

Landslides are among the most common natural hazards and are the most damaging, leading to a variety of human and environmental impacts. The quantitative assessment of landslide hazards for a large area is critical to the mitigation of these loses. Such an assessment is also essential for activities associated with watershed management. This study describes a method for large area landslide hazard assessment, mapping, and evaluation methods and provides an example of a study area from Nepal.

Nepal is a country comprised of 83 percent hills and mountains, and steep terrain, fragile geology, and seasonal monsoon rainfall contribute to the landslide potential. Every year, sediment-related disasters in Nepal result in an average loss of 400 lives and property losses amounting to US $ 17 million (Disaster Prevention Technical Center [DPTC], 1994).

Geographic information systems (GIS) (Burrough, 1986; Aronoff, 1989; Marble, 1990) have overcome many of the difficulties normally associated with the handling of data in the study of geomorphic hazards (Dikau *et al.*, 1996). For example, Walsh *et al.* (1990) and Walsh and Butler (1997) used a GIS technique to illustrate morphometric characteristics of snow-avalanche paths and debris flow. Wadge *et al.* (1993) used a GIS to evaluate risk associated with geomorphic hazards and population vulnerability. Various techniques using GIS for the assessment of landslide hazards have been employed by researchers (e.g., Gupta and Joshi, 1990; Mejia-Navarro *et al.*, 1994; Van Westen, 1994; Brunori *et al.*, 1995; Terlien *et al.*, 1995; Binaghi *et al.*, 1998; Dhakal *et al.*, 1999). Nevertheless, the studies that have used GIS in landslide hazard assessment remain limited, especially for large areas that can take full advantage of GIS.

### General Methods and Techniques for Landslide Hazard Assessment

Among many useful classifications of assessment techniques for landslide hazard, the classifications given by Mantovani *et al.* (1996) and Hartle'n and Viberg (1988) cover the majority. The methods of Mantovani *et al.* are for analytical methods (distribution analysis, qualitative analysis, deterministic analysis, landslide frequency analysis, and statistical analysis). The approach of Hartle'n and Viberg describes hazard types (absolute hazard, monitored hazard, empirical hazard, and relative hazard). Distribution analysis results in a map, which gives information pertaining only to those sites where landslides have occurred in the past. In qualitative analysis, which is also called geomorphological analysis (Kienholz *et al.*, 1984;

A.S. Dhakal is with the Graduate School of Agricultural Sciences, T. Amada is with the Institute of Agricultural and Forestry Engineering, and M. Aniya is with the Institute of Geoscience, all at the University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki, 305-8572, Japan.

A.S. Dhakal is presently with the Department of Forest Resources Management, Faculty of Forestry, University of British Columbia, Vancouver, Canada.

Photogrammetric Engineering & Remote Sensing
Vol. 66, No. 8, August 2000, pp. 981–989.

0099-1112/00/6608–981$3.00/0

McKean *et al.*, 1991), several characteristics of terrain are used to define landforms. The degree of hazard is then evaluated at each site of the terrain based on subjective decision rules. The deterministic approach (e.g., Skempton and Delory, 1957; Okimura, 1982; Mostyn and Fell, 1997) expresses the stability of slopes in terms of the safety factor (absolute hazard). For large areas, the variations in parameters included in the analysis of the safety factor are too large to accurately quantify (Jibson and Keefer, 1989; Mulder, 1991). In landslide frequency analysis (e.g., Capecchi and Focardi, 1988), earthquake and rainfall records are compared with landslide dates to obtain a threshold value for certain frequency levels (monitored hazard). The empirical hazard is assessed from earlier and active landslide data by examining relationships such as those between slope angle and relief (e.g., Zika *et al.*, 1988). The empirical and monitored hazards require continuous, long-term data on the landslides and their causative factors under similar environmental conditions. These data are often unavailable. In statistical analysis, the factors associated with topography, geology, and vegetation which can be considered as indices of the parameters of safety factor are quantified to assess their contributions to landsliding (relative hazard (Yin and Yan, 1988; Wang and Unwin, 1992; Pachauri and Pant, 1992; Sarkar *et al.*, 1995; Mark and Ellen, 1995)). This approach is based on the assumption that future landslides will be more likely to occur under conditions similar to those of previous landslides (Varnes, 1984; Brabb, 1984).

### Approach Employed, Issues, and Objectives

Statistical methods of hazard assessment are particularly appropriate for large areas. The benefit of a statistical model is that landslide assessment can be made rapidly, and site investigation cost is minimized. Moreover, the use of GIS has made this an effective method. A multivariate statistical approach, such as discriminant analysis (Davis, 1986), is considered better than the univariate statistical approach, because the former takes into account the interrelationships between the factors. The need to handle nominal data in discriminant analysis can be overcome by employing Quantification Scaling Type II (Q-S II) analysis (Hayashi, 1952; Hayashi, 1980; Hayashi, 1987), which can incorporate nominal data into the model. Two groups of sample data, the landslide and the non-landslide, are required for discriminant analysis. The critical assumption would be that the sampled data truly represent the population. It is usually stipulated that the two groups be similar in size (Klecka, 1980). Therefore, sampling for a non-landslide group is required due to the very large quantity of non-landslide data as compared to the quantity of landslide data, even if all landslide data are to be utilized. Hence, in hazard assessment based on a small grid-cell, the outcome of the analysis may depend on the sample of landslide and non-landslide data used in the analysis (Aniya, 1985; Van Westen, 1993; Chung *et al.*, 1995; Dhakal *et al.*, 1999).

Selecting a representative landslide group that occupies a very small area is not a difficult task whereas selecting a non-landslide group that occupies a large area is difficult because a 100-km$^2$ area consisting of 25- by 25-m grid cells results in a total of 160,000 grid cells. The use of large grid-cell size (e.g., Cararra, 1983) or a land unit based on the catchment area or slope sections (e.g., Carrara *et al.*, 1991) may result in a hazard map that is overly generalized. The aggregation of data in a unit causes a generalization of the input variables, and the relationship between the landslide and non-landslide group cannot be evaluated at the location of the phenomena themselves (Van Westen, 1993; Chung *et al.*, 1995). Although sampling schemes have been shown to be crucial in the small grid-cell based statistical hazard models, no study exists which has attempted to clarify the problem.

A simple random sampling may be suitable for landslide cells. This approach, however, is not practical for obtaining non-landslide cells, because some parts of the area may be over sampled or under sampled. To overcome this problem for a non-landslide group, either a stratified random or systematic sampling may be effective. GIS is then applied to an examination of the effect of different landslide and non-landslide groups on the outcome of the critical factors and classes, from which hazard maps are produced. With a GIS we can evaluate the classified grid cells at the same location on different hazard maps, a process resulting in what we refer to hereafter as "spatial agreement."

The objectives of this study are (1) to examine the effect of different sample combinations (landslide and non-landslide) in defining the factors and classes contributing to landsliding, (2) to produce landslide hazard maps, and (3) to evaluate the hazard maps produced. To investigate the first objective, GIS and Q-S II analysis were used. The results of Q-S II analysis were then used to produce hazard maps, and spatial agreements in the hazard maps were evaluated using GIS.

## Study Area

The study area is the Kulekhani watershed (124 km$^2$) located in the Lesser Himalayan region of the Himalayan belt in the central region of Nepal. The area lies between 27° 34′ N and 27° 42′ N latitudes and between 85° 01′ E and 85° 12′ E longitudes (Figure 1), with elevations ranging from 1,500 m to 2,600 m. This region of the Himalayan belt in Nepal is highly populated and most prone to landsliding. The average annual rainfall is about 1600 mm. The area is drained by the Palung River, which empties into the Kulekhani reservoir. The reservoir received a tremendous amount of sediments (thirty times the average annual) during the landslide/debris-flow disaster of July 1993 (Dhakal, 1995). This is the only reservoir in Nepal and sup-

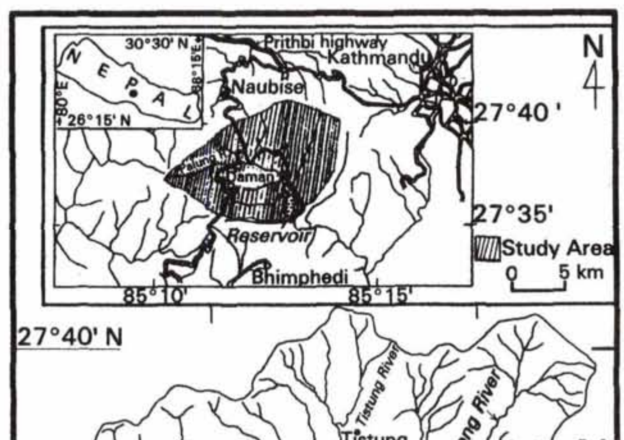


Figure 1. The study area, Kulekhani watershed, Nepal.

ports one third of the total electric power generation of Nepal; consequently, landslide hazard assessment is critical for effective watershed management. The rocks of the study area fall in the Phulchawki group or the Bhimphedi group of the Kathmandu complex of the Lesser Himalaya with granite intrusions (Stocklin and Bhattarai, 1981). The Phulchawki group is characterized by sedimentary or weakly metamorphosed rocks and consists of slates, limestones, and quartzites. The Bhimphedi group contains high-grade meta-sedimentary rocks and consists of slates, meta-sandstones, phyllites, schists, quartzites, and marbles. Crop and forestlands occupy 43 and 44 percent of the total land area, respectively.

## Data Acquisition and GIS Data Layers

Stereopairs of black-and-white vertical aerial photographs (1:20,000 scale) taken in March 1994 were interpreted for landslide identification. Using a stereo zoom-transferscope, landslides identified on the aerial photographs were plotted on a topographic map at a scale of 1:12,500 (photographically enlarged from 1:25,000). The landslide distribution map was finalized and digitized after field verification.

A digital elevation model (DEM) was generated from a triangulated irregular network (TIN) model using digitized contours of topographic maps (contour interval 20 m). The slope gradient was divided by a 10-degree interval into five classes (Table 1). Elevation and slope aspects were divided into four classes each. Ridges and valleys were also defined from the DEM. Employing the Strahler (1957) method for numbering the stream orders, the drainage basin order layer was derived from the topographic map, and divided into three classes. After minor modifications based on fieldwork and aerial photographic verifications, a land-use/land-cover layer was created with five classes from a land-use/land-cover map (1:25,000 scale) produced in 1991 (Department of Forest, Nepal, 1991). The geological map at a scale of 1:50,000 (Nepal Electricity Authority, 1994) was digitized to produce the geology layer. The scale of the geological map is relatively smaller than the other maps used. However, the spatial variation in geology (rock types) relating to landslides is not as fine as other factors such as slope gradient or land use/land cover. The selection and classifications of these factors were primarily guided by the sample landslides surveyed in the field, and previous knowledge of the causal relationships between slope failure and instability factors (e.g., Coates, 1977; Varnes, 1978; Aniya 1985; Crozier, 1986; Zinmermann *et al.*, 1986; Dhakal *et al.*, 1997). Figure 2 depicts the flow of this study.

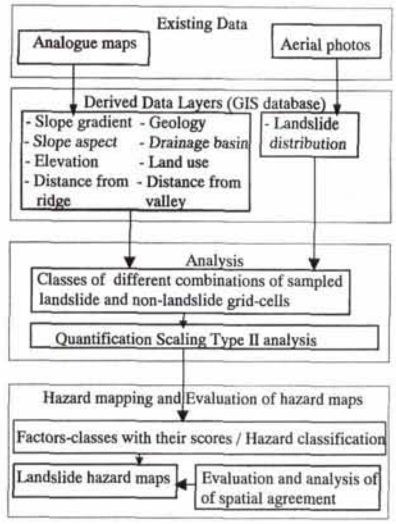


Figure 2. Flow diagram of the study.

## Method of Analysis

### Q-S II Analysis

The Q-S II is a multidimensional quantification analysis (Hayashi, 1950; Hayashi, 1954a) which incorporates nominal data, and is the same as discriminant analysis. The quantification is attained by using frequencies as input data to maximize the efficiency of discrimination (Hayashi, 1952; Hayashi, 1954a). The method is suitable to the landslide hazard assessment, because nominal variables (factors) such as geology or land use/land cover are often most important to discriminate between landslide and non-landslide groups. Other discriminant functions (such as canonical) require interval or ratio data (Klecka, 1983). The linear Q-S II function (score; $\alpha a_q$) for a sample belonging to a group $q$ with $n$ factors and $m$ classes in a factor can be written as

$$\alpha a_q = \sum_{j=1}^{n} \sum_{i=1}^{m} \delta a(ji)\, Xji$$

Table 1. Factors and Their Classes in GIS For a Q-S II Analysis

| Factors | Class Code 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Slope gradient | <15° | 15°–25° | 25°–35° | 35°–45° | >45° | — | — |
| Slope aspect | North (315°–45°) | East (45°–135°) | South (135°–225°) | West (225°–315°) | — | — | — |
| Elevation | <1800 m | 1800 m–2000 m | 2000 m–2200 m | >2200 m | — | — | — |
| Drainage basin order | First | Second | Third | — | — | — | — |
| Distance from ridge | <50 m | 50 m–100 m | >100 m | — | — | — | — |
| Distance from valley | <50 m | 50 m–100 m | >100 m | — | — | — | — |
| Geology | Slates with quartzites or limestones | Limestones | Slates with metasandstones and phyllites | Schists with quartzites and marbles | Biotite schists and micaceous quartzites | Granite | Alluvium |
| Land use/land cover | Crops | Coniferous forest | Broad leaf forest | Mixed forest | Shrub land | — | — |

Table 2. Combinations of Landslide and Non-Landslide Sample Groups for the Analysis

| Combination | Number of non-landslide grid cells | Number of landslide grid cells | Total number of grid cells | Sampling method for non-landslide group | Sampling method for landslide group |
|---|---|---|---|---|---|
| 1 | 572 | 566 | 1138 | unaligned stratified random | Simple random |
| 2 | 571 | 566 | 1137 | " | " |
| 3 | 1143 | 566 | 1709 | " | " |
| 4 | 1285 | 566 | 1851 | Aligned systematic | " |
| 5 | 643 | 566 | 1209 | " | " |

where $\delta a(ji) = 1$ if sample $a$ belongs to the $i$-$th$ class of factor $j$, otherwise 0; and $Xji$ = score of the $i$-$th$ class of factor $j$.

The quantification of classes of the factors ($Xji$) is done in such a way that the proportion of variance between the groups to the total variance (i.e., the correlation ratio, $\eta^2$), which takes the value between zero and one, is maximized. *Eta* ($\eta$) measures the degree of difference between the group means. The efficiency of the discrimination is therefore given by $\eta$ or $\eta^2$ (Hayashi, 1952; Hayashi, 1954a). Because a large class score in a factor contributes more than a small one in the Q-S II functions, a class score and the range of scores of a factor (difference between the maximum and minimum scores of the classes) can be interpreted to determine their importance. For the data, which are not sampled, the factor classes are measured, and the group to which they belong is predicted from the score of the classes. The Q-S II analysis is available in the Japanese version of the SPSS statistical package.

#### Sampling of Landslide and Non-Landslide Groups

Considering the minimum size of the landslides, the study area was tiled into grid cells of 25 m by 25 m, and one set of grid cells (566) representing the landslides was randomly chosen. These grid cells represent about 45 percent of the total number of 1,246 landslides. The remaining 680 landslides (referred to as "test landslide") were later used for the evaluation of the hazard maps produced. Subsequently, the area was stratified into a rectangular block of 2.4 km by 1.5 km, and three sets of non-landslide groups were derived with the same number of grid cells from each block using the unaligned stratified random sampling method. In addition, using the aligned systematic sampling method, two sets of non-landslide groups, with a prior estimation of a sample size, were derived from the same starting value but with a different sampling interval. Altogether, five sets of samples were obtained for the non-landslide group (Table 2). For each grid cell of a set of landslides and five sets of non-landslides, class codes of eight terrain factors (see Table 1) were assigned for the Q-S II analysis.

#### Association between the Factors

A high correlation of factors may simply reflect redundancy without contributing much improvement in the analysis. To address this issue, correlation coefficients between these factors were calculated. Table 3 shows that the results for combination 1 lack strong correlations. The degree of correlation between factors may be different depending on different samples (Liebetrau, 1983). In reality, we cannot expect the causative factors of landslides to be entirely independent. The correlations of factors in five different sample combinations (see Table 2) were examined, and some factors were excluded in order to see their effect on the results of the Q-S II analysis. Based upon field information, the results using all eight factors in the analysis appeared more reasonable in all combinations. Moreover, the maximum separation, as indicated by $\eta$ or $\eta^2$, and the Q-S II accuracy (discussed later) were notably higher when all the eight factors were used.

## Results

The values of $\eta$, $\eta^2$, and the separation between the group centroid (Table 4, bottom) are generally higher for combinations in which the non-landslide group was obtained using the unaligned stratified random sampling method. For the sample size employed in the analysis, the values of $\eta$ are reasonable (Hayashi, 1954b) in all combinations for the discrimination between the landslide and non-landslide groups. The class scores and the range of scores are shown in Table 4. Based upon the range of scores, geology is found to be the most important factor contributing to landsliding in all combinations. The ranking order of the other factors shows some minor variations. Elevation, land use/land cover, and slope aspect fall within the second group of importance, while slope gradient is in the third group. This is true in all combinations except for combination 2, which shows the importance of drainage basin order. Distance from ridge and distance from valley have the least importance.

For the classes of geological factor, the score is high for "granite," followed by "biotite schists with micaceous quartzites." Granite in the study area has been characterized as highly weathered and permeable. With respect to elevation, a zone of "2,000 m–2,200 m" is most susceptible, followed by "1800 m–2000 m." In land use/land cover, "coniferous forest" shows the highest importance followed by "shrub land" in combinations 1 and 3, whereas "shrub land" is followed by "coniferous forest" in combinations 2, 4, and 5. Most of the areas covered by coniferous species are characterized as immature and poorly stocked. As for slope aspect, "East" followed by "South" facing slopes in combinations 1 and 3, and "South" followed by "East" in combinations 2, 4, and 5 are shown to be the most important. This pattern reflects the monsoon rainfall distribu-

Table 3. Correlations Between Factors (Sample of Combination 1)

| | Slope gradient | Slope aspect | Elevation | Drainage basin order | Distance from ridge | Distance from valley | Geology | Land use/ land cover |
|---|---|---|---|---|---|---|---|---|
| Slope gradient | 1.000 | | | | | | | |
| Slope aspect | −0.003 | 1.000 | | | | | | |
| Elevation | 0.153 | 0.014 | 1.000 | | | | | |
| Drainage basin order | 0.165 | 0.068 | 0.224 | 1.000 | | | | |
| Distance from ridge | −0.005 | −0.001 | −0.053 | −0.056 | 1.000 | | | |
| Distance from valley | −0.020 | 0.011 | 0.020 | 0.030 | −0.086 | 1.000 | | |
| Geology | 0.032 | −0.094 | 0.108 | 0.100 | −0.029 | 0.037 | 1.000 | |
| Land use/land cover | 0.085 | −0.036 | 0.128 | 0.113 | −0.033 | 0.012 | 0.131 | 1.000 |

Table 4. The Results of the Q-S II Analysis. A Large Range of Scores Indicates the More Discriminating Potential of the Factor and a Negative Class Score Implies the Contribution to Landsliding

| Factors | Class | Class-code | Combination 1 |  | Combination 2 |  | Combination 3 |  | Combination 4 |  | Combination 5 |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  |  | CS | RS | CS | RS | CS | RS | CS | RS | CS | RS |
| *Slope gradient* | <15° | 1 | *0.138* | *0.573* | *0.074* | *0.268* | *0.109* | *0.438* | *−0.059* | *0.399* | *−0.040* | *0.476* |
|  | 15°–25° | 2 | −0.098 |  | −0.040 |  | −0.088 |  | −0.025 |  | −0.028 |  |
|  | 25°–35° | 3 | 0.025 |  | 0.060 |  | 0.038 |  | 0.141 |  | −0.009 |  |
|  | 35°–45° | 4 | −0.350 |  | −0.194 |  | −0.288 |  | −0.160 |  | −0.079 |  |
|  | >45° | 5 | 0.223 |  | −0.024 |  | 0.150 |  | 0.240 |  | 0.397 |  |
| Slope aspect | North | 1 | 0.396 | 0.775 | 0.473 | 0.831 | 0.484 | 0.834 | 0.343 | 0.832 | 0.173 | 0.862 |
|  | East | 2 | −0.379 |  | −0.214 |  | −0.350 |  | −0.289 |  | −0.251 |  |
|  | South | 3 | −0.215 |  | −0.358 |  | −0.286 |  | −0.412 |  | −0.333 |  |
|  | West | 4 | 0.247 |  | 0.128 |  | 0.149 |  | 0.420 |  | 0.529 |  |
| Elevation | <1800 m | 1 | 0.283 | 0.707 | 0.335 | 0.835 | 0.318 | 0.856 | 0.506 | 1.055 | 0.317 | 0.705 |
|  | 1800 m–2000 m | 2 | −0.103 |  | −0.177 |  | −0.134 |  | −0.063 |  | −0.006 |  |
|  | 2000 m–2200 m | 3 | −0.365 |  | −0.412 |  | −0.485 |  | −0.548 |  | −0.389 |  |
|  | >2200 m | 4 | 0.342 |  | 0.423 |  | 0.371 |  | 0.045 |  | 0.173 |  |
| Drainage basin order | First | 1 | −0.023 | 0.143 | −0.078 | 0.502 | −0.054 | 0.332 | 0.028 | 0.204 | 0.010 | 0.267 |
|  | Second | 2 | −0.003 |  | −0.043 |  | −0.033 |  | −0.137 |  | −0.126 |  |
|  | Third | 3 | 0.120 |  | 0.424 |  | 0.278 |  | 0.067 |  | 0.141 |  |
| Distance from ridge | <50 m | 1 | 0.067 | 0.135 | 0.002 | 0.081 | 0.043 | 0.106 | 0.090 | 0.247 | −0.007 | 0.249 |
|  | 50 m–100 m | 2 | 0.028 |  | 0.048 |  | 0.045 |  | 0.103 |  | 0.147 |  |
|  | >100 m | 3 | −0.068 |  | −0.033 |  | −0.061 |  | −0.144 |  | −0.102 |  |
| Distance from Valley | <50 m | 1 | −0.090 | 0.156 | −0.097 | 0.145 | −0.096 | 0.137 | −0.069 | 0.173 | −0.135 | 0.199 |
|  | 50 m–100 m | 2 | −0.010 |  | 0.033 |  | 0.036 |  | 0.104 |  | 0.064 |  |
|  | >100 m | 3 | 0.066 |  | 0.048 |  | 0.041 |  | −0.029 |  | 0.051 |  |
| Geology | Slates with quartzites or limestones | 1 | 1.883 | 2.472 | 1.849 | 2.557 | 1.607 | 2.333 | 1.657 | 2.364 | 1.848 | 2.415 |
|  | Limestones | 2 | 0.630 |  | 0.894 |  | 0.818 |  | 0.771 |  | 0.717 |  |
|  | Slates with meta-sandstones and phyllites | 3 | 0.075 |  | 0.169 |  | 0.112 |  | 0.109 |  | 0.121 |  |
|  | *Schists with quartzites and marble* | 4 | *0.195* |  | *0.069* |  | *0.110* |  | *0.162* |  | *0.101* |  |
|  | Biotite schists and micaceous quartzites | 5 | −0.266 |  | −0.099 |  | −0.210 |  | −0.311 |  | −0.321 |  |
|  | Granite | 6 | −0.589 |  | −0.708 |  | −0.726 |  | −0.707 |  | −0.567 |  |
|  | Alluvium | 7 | 1.359 |  | 1.370 |  | 0.888 |  | 1.097 |  | 1.454 |  |
| Land use/land cover | Crops | 1 | 0.368 | 0.804 | 0.326 | 0.769 | 0.348 | 0.789 | 0.308 | 0.824 | 0.421 | 0.881 |
|  | Coniferous forest | 2 | −0.436 |  | −0.327 |  | −0.441 |  | −0.315 |  | −0.393 |  |
|  | Broad leaf forest | 3 | 0.137 |  | 0.219 |  | 0.179 |  | 0.140 |  | 0.0094 |  |
|  | Mixed Forest | 4 | −0.175 |  | −0.136 |  | −0.193 |  | −0.037 |  | −0.111 |  |
|  | Shrub land | 5 | −0.262 |  | −0.443 |  | −0.395 |  | −0.515 |  | −0.461 |  |
| $\eta$ |  |  | 0.450 |  | 0.460 |  | 0.412 |  | 0.379 |  | 0.423 |  |
| $\eta^2$ |  |  | 0.203 |  | 0.211 |  | 0.170 |  | 0.143 |  | 0.179 |  |
| Group centroid: |  |  |  |  |  |  |  |  |  |  |  |  |
| Landside group |  |  | −0.452 |  | −0.462 |  | −0.585 |  | −0.571 |  | −0.356 |  |
| Non-landslide group |  |  | 0.448 |  | 0.457 |  | 0.289 |  | 0.251 |  | 0.314 |  |

CS = Class score, RS = Range of scores

tion, which contributes 80 percent of the total annual rainfall, giving southern and eastern faces more rainfall. For slope gradient, the score is highest at class "35°–45°." The first-order drainage basin is found most susceptible in combinations 1, 2, and 3 while combinations 4 and 5 show higher importance of the second order drainage basin. The greater the distance from the ridge (>100 m) and the shorter the distance from the valley (<50 m), the greater the susceptibility to landslides.

In summary, the most critical association of classes for landsliding are "granite," "2000 m–2200 m," "coniferous forest" ("shrub land" in combinations 2, 4, and 5), "East" ("South" in combinations 2, 4, and 5), "35°–45°," "first order" ("second order" in combinations 4 and 5), <50 m (distance from valley), and >100 m (distance from ridge).

A similar study conducted for the Amahata River Basin in Japan by Aniya (1985) found that vegetation, slope gradient, and slope aspect were the factors critical to landsliding. A study in Omichi-Dani, Japan by Amada *et al.* (1995) found geology and slope aspect as the important factors causing landslides. Dikau (1990) found slope angle and concave form elements (plan and profile) as the significant descriptors of the landslides in the Mainz Basin, southwest Germany. It is to be expected that different regions have different critical factors, because the environmental factors such as geology, land use, and topography also are likely to be different as well as the climatic and hydrological conditions.

## Classification of Hazard Classes and Mapping

The factor layers with scores for their classes were superimposed to obtain a cumulative score at each grid cell. The following discussion illustrates the classification of grid cells into different hazard classes. Figure 3 shows the plots of cumulative frequencies for the sample data of combination 1, with the landslide group starting from the smallest score and the non-landslide group from the largest score. If the curves of these two groups do not intersect, then separation is perfect. However, in this case, the curves do intersect and misclassifications will occur. For a boundary score that would separate the two groups, it appears logical to choose the score at which the two curves intersect, i.e., −0.13 (discrimination score) in this case, because it is as important to correctly locate stable slopes as unstable ones. Consequently, grid cells whose total score of the classes was equal to or less than −0.13 were classified into the landslide group (unstable), whereas those with the score greater than −0.13 were placed in the non-landslide group (stable). Then the overall accuracy (Q-S II accuracy) for the sample

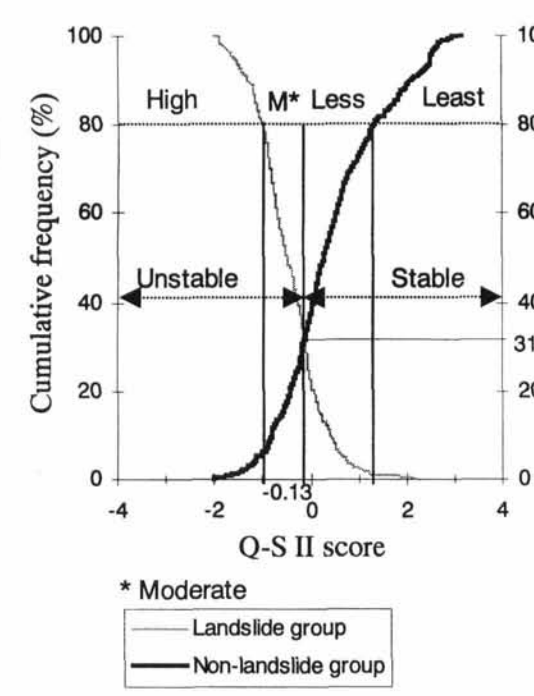


Figure 3. Cumulative frequencies and Q-S II scores for landslide and non-landslide groups (Combination 1; see caption of Figure 4 for explanation of Combination 1).

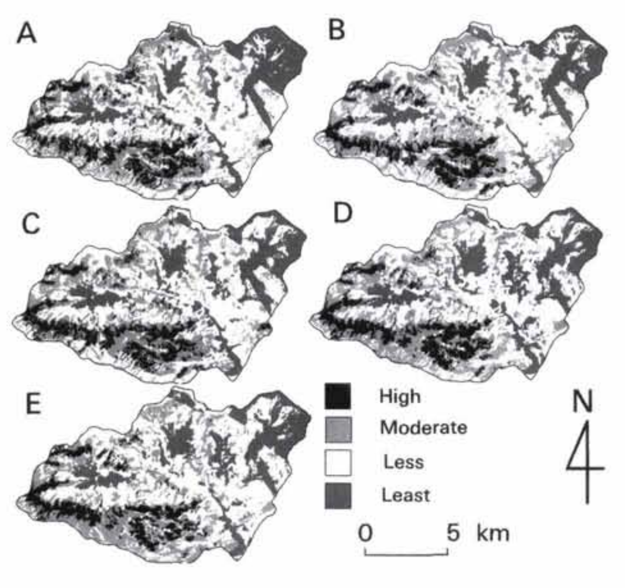


Figure 4. Landslide hazard maps produced from the results of different sample combinations employed in the Q-S II analysis. (A) Combination 1 (566 randomly sampled landslide grid cells and 572 unaligned stratified randomly sampled non-landslide grid cells). (B) Combination 2 (566 randomly sampled landslide grid cells and 571 unaligned stratified randomly sampled non-landslide grid cells). (C) Combination 3 (566 randomly sampled landslide grid cells and 1143 unaligned stratified randomly sampled non-landslide grid cells). (D) Combination 4 (566 randomly sampled landslide grid cells and 1285 aligned systematically sampled non-landslide grid cells). (E) Combination 5 (566 randomly sampled landslide grid cells and 643 aligned systematically sampled non-landslide grid cells).

data of combination 1 is 68.5 percent (100 − 31.5; Figure 3). The Q-S II accuracy for combinations 2, 3, 4, and 5 are 70.1 percent, 69.2 percent, 67.5 percent, and 67.3 percent, respectively. In order to differentiate between the very unstable and marginally unstable categories, and the very stable and marginally stable categories, scores at which the accuracy of decision would be about 80 percent were selected. This resulted in the division of each category into two classes, resulting in four classes of relative hazard: high, moderate, low, and least. Figure 4 depicts five hazard maps produced from the results of the five sample combinations (see Table 2). Table 5 compares the percentage area of hazard classes in five hazard maps in which hazard classes do not show substantial differences.

## Evaluation of Hazard Maps

In order to determine which sample combination best represents the population, the accuracy of hazard maps (i.e., evaluation of Q-S II results) was assessed. In addition, the spatial agreements between the hazard maps were measured to comprehensively examine the effect of sampling on the final outcome of the analysis.

### Accuracy of Hazard Maps

A common method for the evaluation of landslide hazard map is to compute the percentage of landslide for each hazard class (Van Westen, 1993; Dhakal *et al.*, 1999). A large number of landslide grid cells in the unstable area should indicate higher accuracy of the hazard map. Grid cells that lack landslides but are classified as unstable may indicate that they are potentially unstable (Neuland 1976; Carrara, 1983). The percentage of "test landslide" grid cells (other than sampled landslides for analysis) in the unstable category is a measure of accuracy of the hazard maps. Then, for the evaluation of hazard maps, the percentage of "test landslide" grid cells in the unstable category can be compared to the Q-S II accuracy of the sampled landslide data (see Figure 3).

Combinations 4 and 5 are those in which non-landslide groups were obtained by using an aligned systematic sampling method. The accuracies for these two combinations are comparatively lower than for the other combinations (Table 6). In the aligned systematic sampling method, the selection of the starting grid cells predetermined the position of all subsequent grid cells. In this study, 196 non-landslide group grid cells in Combination 4 were found partially or fully existing within the 50-m range of landslides. This number is only three for Combination 1. These suggest the periodicity of the landslide distribution. In the aligned systematic sampling method, a large

TABLE 5. PERCENTAGE AREA OF DIFFERENT HAZARD CLASSES IN FIVE HAZARD MAPS

| | Percentage area of different hazard classes (%) | | | |
|---|---|---|---|---|
| Combination | High | Moderate | Less | Least |
| Combination 1 | 7.7 | 30.9 | 44.1 | 17.3 |
| Combination 2 | 7.8 | 30.5 | 44.1 | 17.6 |
| Combination 3 | 7.3 | 30.6 | 44.8 | 17.3 |
| Combination 4 | 8.3 | 29.0 | 43.7 | 19.0 |
| Combination 5 | 8.0 | 29.0 | 44.5 | 18.5 |

TABLE 6. TEST LANDSLIDE GRID CELLS FALLING ON UNSTABLE CATEGORY. FOR COMPARISION Q-S II ACCURACY IS ALSO LISTED

| Combination | Test landslide in unstable category (%) | Q-S II accuracy (%) |
|---|---|---|
| Combination 1 | 67.7 | 68.5 |
| Combination 2 | 65.9 | 70.1 |
| Combination 3 | 67.7 | 69.2 |
| Combination 4 | 65.7 | 67.5 |
| Combination 5 | 62.8 | 67.3 |

TABLE 7. COMPARISON OF SPATIAL AGREEMENT (%) BETWEEN FIVE HAZARD MAPS

| | Combination 1 | Combination 2 | Combination 3 | Combination 4 | Combination 5 |
|---|---|---|---|---|---|
| Combination 1 | 100 | | | | |
| Combination 2 | 82.5 | 100 | | | |
| Combination 3 | 88.8 | 90.1 | 100 | | |
| Combination 4 | 79.6 | 81.1 | 82.7 | 100 | |
| Combination 5 | 79 | 77.9 | 79.2 | 87.7 | 100 |

number of non-landslide grid cells adjacent to or near the landslides might have given those grid cell site characteristics similar to landslides, due to the effect of autocorrelation, thereby lowering the accuracy. Combination 2 has the highest accuracy of landslide identification in the Q-S II analysis (sample data). *This combination resulted in a comparatively low accuracy of* hazard map compared to combinations 1 and 3 (see Table 6), although not significant.

#### Evaluation of Spatial Agreement

To evaluate the spatial agreement of the hazard classes between the hazard maps, two of the five hazard maps were overlaid in turn, and all the grid cells classified into the same hazard class (agreed grid cells) were counted. The "overall spatial agreement" was then calculated by taking the proportion of agreed grid cells to the total number of grid cells, in a manner similar to the evaluation of overall accuracy from the error matrix (e.g., Congalton *et al.*, 1983; Congalton, 1991). Table 7 compares the overall spatial agreement between the five hazard maps. Figure 5 is a visual example, which depicts the agreed and disagreed hazard classifications when the hazard map of combination 1 was crossed with the remaining four. The disagreement in the hazard maps varies between 10 and 20 percent. The agreement between the hazard maps is higher for the hazard maps resulting from sample combinations that used the same method for selecting a non-landslide group (see Table 2 and Table 7). It is also important to point out that disagreement between the hazard maps was introduced only from the immediate class, i.e., *no "high" hazard class identified in a hazard map was classified as* "less" or "least" in the other and vice versa.

## Conclusions

Landslide hazard assessment involves many different problems at various stages of analysis. A GIS can facilitate a trial-and-error approach for assessment methods. The sampling technique is one of the most important aspects of hazard analysis. We have shown in this study that the use of a GIS is invaluable for determining the best sampling techniques for the grid cell based hazard assessment. The factors considered may also be dependent upon the data available because creating new data is time consuming and costly, if not impossible. Also, local site characteristics are often difficult to incorporate into hazard *assessment for large areas.*

Our results suggest that the unaligned stratified random sampling method is better than the aligned systematic sampling method for selecting a non-landslide group, because the former yielded the highest QS-II indices, and resulted in the most accurate hazard map. A small variation in the overall spatial agreements can be recognized in the three hazard maps, which used the results from the unaligned stratified random samples

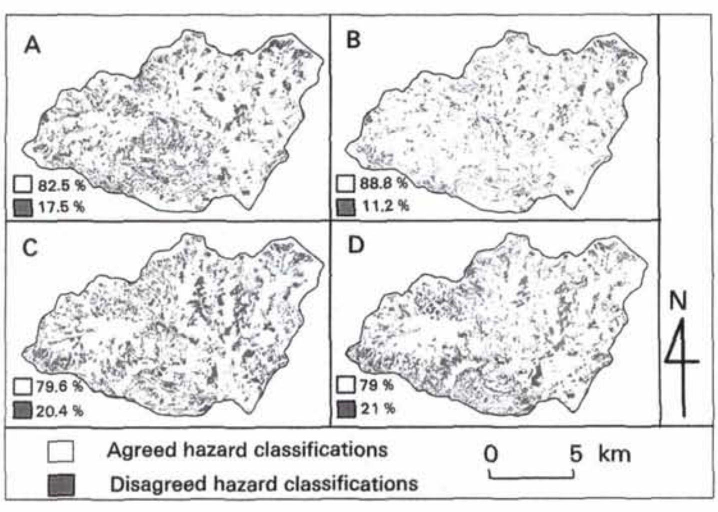


Figure 5. Spatial agreement and disagreement in the hazard classes between different hazard maps. The figure depicts the effect of sampling schemes on the hazard maps (see captions of Figure 4 for explanations of combinations). (A) Combination 2 and Combination 1. (B) Combination 3 and Combination 1. (C) Combination 4 and Combination 1. (D) Combination 5 and Combination 1.

of non-landslide group. These results are likely to indicate that, although the outcome depends to some extent on the sampled data, the difference may be small enough for the practical use of a hazard map. As a result, when using unaligned stratified random sampling for a non-landslide group, one sample set would practically suffice for the analysis and hazard mapping.

## Acknowledgments

The logistic support provided during the fieldwork by the Disaster Prevention Technical Center (DPTC)/JICA, Kathmandu, is highly acknowledged. The research grant awarded by the Japan Society of Erosion Control Engineering in the year 1998 is acknowledged. Mr. Soonchan Park (Ph.D. Candidate, Institute of Geoscience, University of Tsukuba) is thanked for his contributions at various stages of the work. The authors gratefully acknowledge David Sprague, Ph.D. (Senior Researcher, National Institute for Agro-Environmental Sciences, Tsukuba, Japan) for his valuable comments on the manuscript. Ogbonna James Chukwuma (Associate Professor, University of Tsukuba) and Stephen Bird (English Teacher, Meikei High School, Tsukuba) deserve special thanks for their editorial comments. The suggestions made by three anonymous reviewers are highly appreciated.

## References

Amada, T., M. Aniya, and T. Matsubara, 1995. GIS application to landslide susceptibility mapping—As a first step toward modeling sediment discharge, *Proceedings of the International Sabo Symposium*, Tokyo, pp. 291–298.

Aniya, M., 1985. Landslide-susceptibility mapping in the Amahata River Basin, Japan, *Annals of the Association of American Geographers*, 75(1):102–114.

Aronoff, S., 1989. *Geographic Information System: A Management Perspective*, WDL Publications, Ottawa, 279 p.

Binaghi E., L. Luzi, P. Madella, F. Pergalani, and A. Rampini, 1998. Slope instability zonation: A comparison between certainty factor and fuzzy dempster-shafer approaches, *Natural Hazards*, 17:77–97.

Brabb, E.E., 1984. Innovative approach to landslide hazard and risk mapping, *Proceedings of the 4th International Symposium on Landslides*, Toronto, Canada, 1:307–324.

Brunori, F., F. Casagli, S. Flaschi, C.A. Garzonio, and S. Moreti, 1995. Landslide hazard mapping in Tuscany, Italy: An example of automatic evaluation, *Geomorphic Hazards* (O. Slaymaker, editor), John Wiley and Sons, West Sussex, pp. 55–67.

Burrough, P.A., 1986. *Principles of Geographical Information Systems for Land Resources Assessment*, Clarendon Press, Oxford, 194 p.

Capecchi, F., and P. Focardi, 1988. Rainfall and landslides: Research into a critical precipitation coefficient in an area of Italy, *Proceedings of the 5th International Symposium on Landslides*, Lausanne, 2:1031–1136.

Carrara, A., 1983. Multivariate models for landslide hazard evaluation, *Mathematical Geology*, 15(3):403–427.

Carrara, A., M. Cardinali, R. Detti, F. Guzzetti, N. Pasqui, and P. Reichenbach, 1991. GIS technique and statistical models in evaluating landslide hazard, *Earth Surface Process and Land Forms*, 16:427–445.

Chung, C.F., A.G. Fabbri, and C.J. Van Westen, 1995. Multivariate regression analysis for landslide hazard zonation, *Geographical Information Systems in Assessing Natural Hazards* (A. Carrara and F. Guzzetti, editors), Kluwer Academic Publishers, Dordrecht, pp. 107–134.

Coates, D.R., 1977. Landslide perspectives, *Landslides* (D.R. Coates, editor), Geological Society of America, Boulder, Colorado, pp. 3–28.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37:35–46.

Congalton, R.G., R.G. Oderwald, and R.A. Mead, 1983. Assessing Landsat classification accuracy using discrete multivariate statistical techniques, *Photogrammetric Engineering & Remote sensing*, 49(12):1671–1678.

Crozier, J.M., 1986. *Landslides: Causes, Consequences and Environment*, Croom Helm Ltd., London, 252 p.

Davis, J.C., 1986. *Statistics and Data Analysis in Geology*, John Wiley & Sons, New York, 646 p.

Department of Forest, Nepal, 1991. *Land Use Map of Kulekhani Watershed* (1:25,000), Finnish International Dev. Agency (FINNIDA)/ Department of Forest, 1 sheet.

Dhakal, A.S., 1995. Disaster of 1993 experienced by Nepal, *Proceedings of the International Sabo Symposium*, Tokyo, pp. 115–122.

Dhakal, A.S., T. Amada, and M. Aniya, 1997. A GIS approach to landslide susceptibility mapping: A case study from the Kulekhani watershed, Nepal, *Proceedings of the 18th Asian Conference on Remote Sensing*, Kuala Lumpur, JS-6-1-6-6.

———, 1999. Landslide hazard mapping and the application of GIS in the Kulekhani watershed Nepal, *Mountain Research and Development*, 19(1):3–16.

Dikau, R., 1990. Derivatives from detailed geoscientific maps using computer methods, *Z. Geomorph* N.F., Suppl. Bd, 80:45–50.

Dikau, R., A. Cavallin, and S. Jager, 1996. Databases and GIS for landslide research in Europe, *Geomorphology*, 15:227–239.

Disaster Prevention Technical Center (DPTC), 1994. *Annual Disaster Review 1993*, Water Induced Disaster Prevention Technical Center, Lalitpur, Nepal, 90 p.

Gupta, R.P., and B.C. Joshi, 1990. Landslide hazard zoning using the GIS approach—A case study from the Ramganga catchment, Himalayas, *Engineering Geology*, 28:119–131.

Hartle'n, J., and L. Viberg, 1988. General report: Evaluation of landslide hazard, *Proceedings of the 5th International Symposium on Landslides*, Lausanne, 2:1037–1057.

Hayashi, C., 1950. On the quantification of qualitative data from the mathematico-statistical point of view, *Annals of the Institute of Statistical Mathematics*, 1:35–47.

———, 1952. On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view, *Annals of the Institute of Statistical Mathematics*, 3:69–98.

———, 1954a. Multidimensional quantification—with the applications to analysis of social phenomena, *Annals of the Institute of Statistical Mathematics*, 5:121–143.

———, 1954b. Examples of theory of Quantification, *Proceedings of the Institute of Statistical Mathematics*, Tokyo, 2(1):11–30 (in Japanese with English abstract).

———, 1980. Quantification of qualitative Data—statistical analysis of categorical data, *Analyse Des Donnees, Actes de la Journee de Travail*, Institut National De Recherche En Informatique Et En Automatique, Naples, pp. 271–281.

———, 1987. *New Developments in Multidimensional Data Analysis*, Recent Developments in Clustering and Data Analysis (C. Hayashi, E. Diday, M. Jambu, and N. Oshumi, editors), Proceedings of the Japanese-French Scientific Seminar, Academic Press, Inc., San Diego, California, pp. 3–16.

Jibson, R.W., and D.K. Keefer, 1989. Statistical analysis of factors affecting landslide distribution in the New Madrid seismic zone, Tennessee and Kentucky, *Engineering Geology*, 27:509–542.

Kienholz, H., G. Schneider, M. Bichsel, M. Grunder, and P. Mool, 1984. Mapping of mountain hazards and slope stability, *Mountain Research and Development*, 4(3):247–266.

Klecka, R.W., 1980. *Discriminant Analysis*, Sage University Paper Series Quantitative Applications in the Social Sciences, Series No. 07-019, Sage Publications, Beverly Hills and London, 71 p.

Liebetrau, A.M., 1983. *Measures of Association*, Sage University Paper Series Quantitative Applications in the Social Sciences, Series No. 07-032, Sage Publications, Beverly Hills and London, 95 p.

Mantovani, F., R. Soeters, and C.J. Van Westen, 1996. Remote sensing techniques for landslide studies and hazard zonation in Europe, *Geomorphology*, 15:213–225.

Marble, D.F., 1990. *Geographic information systems: An overview, Introductory Readings in Geographic Information Systems* (D.J.

Peuquet and D.F. Marble, editors), Taylor and Francis, London, pp. 8–17.

Mark, R.K., and S.D. Ellen, 1995. Statistical and simulation models for mapping debris-flow hazard, *Geographical Information Systems in Assessing Natural Hazards* (A. Carrara and F. Guzzetti, editors), Kluwer Academic Publishers, Dordrecht, pp. 93–106.

McKean, J., S. Buechel, and L. Gaydos, 1991. Remote sensing and landslide hazard assessment, *Photogrammetric Engineering & Remote sensing*, 57(9):1185–1193.

Mejia-Navarro, M., E.E. Wohl, and S.D. Oaks, 1994. Geological hazards, vulnerability, and risk assessment using GIS: Model for Glennwood Springs, Colorado, *Proceedings of the 25th Binghampton Symposium in Geomorphology* (M. Morisawa, editor), Elsevier, Amsterdam, pp. 331–354.

Mostyn, G.R, and R. Fell, 1997. Quantitative and semiquantitative estimation of the probability of landsliding, *Landslide Risk Assessment* (D. Cruden and R. Fell, editors), A.A. Balkema, Rotterdam, Brookfield, pp. 297–315.

Mulder, H.F.H.M., 1991. *Assessment of Landslide Hazard*, Faculty of Geographical Sciences, University of Utrecht, The Netherlands, 150 p.

Nepal Electricity Authority, 1994. *Master Plan Study on Sediment Control for Kulekhani Watershed*, Kulekhani Disaster Prevention Project, Nepal Electricity Authority, Kathmandu, 193 p.

Neuland, H., 1976. A prediction model for landslips, *Catena* 3:215–230.

Okimura, T., 1982. Situation of surficial slope failure based on the distribution of soil layer, *Shin-sabo*, 35:9–18 (in Japanese with English abstract).

Pachauri, A.K., and M. Pant, 1992. Landslide hazard mapping based on geological attributes, *Engineering Geology*, 32:81–100.

Sarkar, S., D.P. Kanungo, and G.S. Mehotra, 1995. Landslide hazard zonation, a case study in Garhwal Himalaya, India, *Mountain Research and Development* 15(4):301–309.

Skempton, A.W., and F.A. Delory 1957. Stability of natural slopes in London clay, *Proceedings of 4th International Conference on Soil Mechanics and Foundation Engineering*, 4:379–381.

Stocklin, J., and K.D. Bhattarai, 1981. *Geology of Kathmandu Area and Central Mahabharat Range, Nepal Himalaya*, United Nations Mineral Exploration Technical Report, UNDP/Nepal, 61 p.

Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology, *Transactions of the American Geophysical Union*, 38:913–20.

Terlien, M.T.J., C.J. Van Westen, and T.W.J. Van Asch, 1995. Deterministic modelling in GIS-based landslide hazard assessment, *Geographical Information Systems in Assessing Natural Hazards* (A. Carrara and F. Guzzetti, editors), Kluwer Academic Publishers, Dordrecht, pp. 57–78.

Van Westen, C.J., 1993. *Application of Geographic Information System to Landslide Hazard Zonation*, ITC Publication Number 15, Part 1, ITC, Enschede, 245 p.

———, 1994. GIS in landslide hazard zonation: A review, with examples from the Columbian Andes, *Mountain Environments and Geographic Information Systems* (M.F. Price and D.I. Heywood, editors), Taylor and Francis, London, pp. 135–166.

Varnes, D.J., 1978. Slope movement types and processes, *Landslide, Analysis and Control* (R.L. Schuster and R.J. Krizek, editors), Transportation Research Board, National Research Council, National Academy of Sciences, Washington, D.C., pp. 11–33.

———, 1984. *Landslide Hazard Zonation: A Review of Principles and Practice*, Commission on the Landslides of the IAEG, UNESCO, Natural Hazard No. 3, 61 p.

Wadge, G., A.P. Wislocki, and E.J. Pearson, 1993. Spatial analysis in GIS for natural hazard assessment, *Environmental Modelling with GIS* (M.F. Goodchild, B.O. Parks, and L.T. Steyaert, editors), Oxford University Press, Oxford, pp. 332–338.

Walsh, S.J., D.R. Butler, E.G. Brown, and L. Bian, 1990. Cartographic modeling of snow avalanche path locations in Glacier National park, Montana, *Photogrammetric Engineering & Remote Sensing*, 56(5):615–621.

Walsh, S.J., and D.R. Butler, 1997. Morphometric and multispectral image analysis of debris flows for natural hazard assessment, *Geocarto International*, 12(1):59–70.

Wang, S.Q., and D.J. Unwin, 1992. Modeling landslide distribution on loess soils in China: An investigation, *International Journal of Geographical Information Systems*, 6:391–405.

Yin, K.L., and T.Z. Yan, 1988. Statistical prediction model for slope instability of metamorphosed rocks, *Proceedings of the 5th International Symposium on Landslides*, Lausanne, 2:1269–1272.

Zika, P., J. Rubar, and Z. Kudrna, 1988. Empirical approach to the evaluation of the stability on high slopes, *Proceedings of the 5th International Symposium on Landslides*, Lausanne, 2:1273–1275.

Zinmermann, M., M. Bichsel, and H. Kienholz, 1986. Mountain hazard mapping in the Khumbu Himal, Nepal with prototype map, scale 1:50,000, *Mountain Research and Development*, 6(1): 29–40.

(Received 19 January 1999; revised and accepted 07 September 1999)