

MODELING THE CONDITIONAL PROBABILITY OF THE OCCURRENCES OF FUTURE LANDSLIDES IN A STUDY AREA CHARACTERIZED BY SPATIAL DATA

Chang-Jo Chung^{a*} and Andrea G. Fabbri^b

a: Geological Survey of Canada, 601 Booth Street, Ottawa, Canada K1A 0E8
E-mail: chung@gsc.nrcan.gc.ca

b: ITC, Hengelosestraat 99, 7500 AA Enschede, The Netherlands
E-mail: fabbri@itc.nl

Commission IV, WG IV/1

KEY WORDS: Landslide hazard, conditional probability, cross-validation, likelihood ratio function

ABSTRACT:

The most crucial but difficult task in the analysis of the risk due to landslide hazard is the estimation of the conditional probability of the occurrence of future landslides in a study area within a specific time period given the presence of spatial and geomorphologic features. This contribution explores a modeling procedure for estimating that conditional probability. The procedure proposed consists of two steps. The first step is to divide the study area into a number of "prediction" classes according to the hazard level for the likely occurrence of future landslides. "Favourability Functions" based on the spatial and geomorphological data in the study area were used for the sub-division. The number of the classes is dependent on the quantity and quality of the input data. Each class represents a level of hazard with respect to the future landslides. We term it the "hazard-mapping step". For this step, several quantitative models have been developed and the strategy is to reconstruct the typical settings in which the future landslides are likely to occur. The second step is to empirically estimate the conditional probability in each prediction class given the spatial and geomorphologic data based on cross-validation techniques. For the second step, termed the "probability estimation step" the basic strategy of the cross-validation is to construct the prediction classes in the first step using the occurrences of the landslides from the first time-period and then to compare the prediction classes with the distribution of the landslide occurrences from the later time period. The statistics obtained from the comparison provides the crucial quantitative measure to estimate the conditional probability. We illustrate the modeling procedure using a case study, La Baie, Quebec in Canada.

1. Introduction

For a given study area, geomorphologists, experts in surficial earth processes have traditionally constructed a landslide hazard map identifying areas likely to be affected by future landslides. It has been achieved by geomorphological understanding of the area through aerial photographs and field works. The hazard map is usually derived from geomorphological maps containing the basic geomorphological characteristics of landforms and it includes a systematic inventory of the past landslides (Panizza et al, 1998). On the other hand, quantitative geomorphologists and civil engineers have constructed a slope stability map based on deterministic models by studying and interpreting the physical processes of landslides using slope angles, soil cohesion, water saturation capacities, shearing resistance and etc. Each point in the stability map shows a level for "the safety factor of slope failure" of the unit area surrounded the point (Terlien et al, 1995).

While the hazard maps from geomorphological maps usually show three or five levels of hazard, the slope stability maps are shown the level for the safety factor in a continuous scale. These prediction maps representing both the hazard and the slope stability maps are generated for guiding the decision makers for land-use planning. The difficulty facing the land-use planners is how to interpret the hazard levels. For example, if only a small sub-area has been assigned as "extreme hazard class" in a hazard map or has

consistently extreme values for the safety factors of slope failures in slope stability map, then it may be relatively easy to make a decision not to allow any types of economic and human activities, the economic sterilization by the ban may outweigh the possible future damage due to the occurrences of future landslides in that small sub-area. However, if a sub-area is classified as "high hazard class" or has a high value for the safety factor in a relatively large sub-area, then although it obviously indicates that the sub-area is possibly be affected by a future landslide, the decision of what to do with the sub-area becomes much more difficult, because the decision makers must compare the economic sterilization with the possible damage.

What the decision makers want to have from the hazard maps or the slope stability maps is not only the relative levels of hazard but also the estimates of the probabilities of the occurrences of future landslides in any given points under certain future scenarios such as a number of future landslides are going to be occurred in the study area within the next 30 years. If we have such estimates of the probabilities, then based on a cost-benefit analysis the decision makers can quantitatively compare the economic sterilization with the possible damage under the assumptions of the scenarios, and hence make a learned and an informed decision, rather than an emotional or a "gut-feeling" decision.

We have adapted a two-step approach proposed by Chung (2002) to tackle the problem of estimating the

probability of the occurrence of future landslide given geomorphological information on a sub-area under an assumption of a scenario. The first step is to construct a hazard map with a number of hazard classes as similarly done by the geomorphologists or civil engineers. Then the second step is to estimate the probability in each class given a scenario. Depending on the availability of the input data set with respect to the locations and timings of the past landslides, it may not be possible to estimate the probabilities of the occurrences of future landslides under certain scenarios (Chung, 2002).

The basic strategy is that the occurrences of the past landslides in the study area are first divided into mutually exclusive two groups. One group is used to build a prediction model, which will generate a prediction map showing several levels of prediction classes. Counting the landslides in the other group in the prediction classes of the prediction map, we estimate the conditional probabilities of the occurrences of the future landslides in the prediction classes.

Consider a study area with m layers of spatial data. Each layer consists of several non-overlapping thematic classes such as soil types or observations of a continuous measurement such as slope angle, which represent coverage of “one theme” known to correlate spatially and genetically with the type of landslides under study. At each pixel in the study area, m values are observed, one value representing the thematic classification of the pixel or a continuous measurement at the pixel for each layer. Using these m values at the pixel, we wish to construct a prediction model, which measures the hazard of future landslides at the pixel.

Several favourability function models (Chung and Fabbri, 1993, 1998, 1999, 2001, 2002) have been developed to tackle the prediction models. To illustrate the proposed strategy, we have selected the likelihood ratio function model as the prediction model. We have used a case study from La Baie area in Quebec, Canada (Chung and Perret, 2002).

2. Input spatial data matrix.

Let \mathcal{A} denote the study area. Consider that we have m map-layers (causal factors) containing geomorphological information related to the occurrences of landslides in \mathcal{A} . Each layer contains one particular causal factor such as surficial geological information or the slope angles. In addition to the m map-layers, we also have the landslide map-layer containing the scars of the past landslides in \mathcal{A} . When the landslide scars are rather small with respect to the map scale, we may have only the locations of the past landslides as points on the map rather than the polygons of the scars. In each scar, geomorphologists typically delineate the scarp where the landslide is triggered. Whenever we refer a landslide in this manuscript, we mean it as either the scarp or the point location of the landslide.

For a quantitative study, we overlay a fine grid over \mathcal{A} , such that each grid cell covers a small unit area. The size of the unit area is depended on both the original input-map-scales and the purpose of the study. The size ranges typically from 5m x 5m to 50m x 50m on ground. Each grid cell is termed a pixel. For each map-layer, one data value, termed pixel value is assigned to each pixel. Consider the case study of La Baie, Quebec. The database typically contains the binary information on the past landslides, a

number of thematic classification maps such as the bedrock geology, and DEM information. From the database, a data matrix is constructed for quantitative analysis. In the data matrix, for the landslide map-layer, Y , “1” is assigned to the pixel value, when more than 50% of a pixel is covered by a scarp of a past landslide. Otherwise, “0” is assigned to the pixel value.

A digital elevation model (DEM) containing three spatial information at each pixel: (i) slope angle; (ii) aspect angle; and (iii) elevation, was included in the input data matrix. From the elevation contours, we usually obtain DEM.

The input data matrix usually consists of two different types of spatial data: (i) categorical data layers such as geological map containing rock types, and (ii) continuous data layers such as slope angles. Combining these two different types of data layers is one of the difficulties of constructing prediction maps.

La Baie, Quebec Study area covers 10km x 6km. The image consists of 2000 x 1200 pixels and each pixel covers 5m x 5m in ground. We have five layers of geomorphological information related to the landslides in the area. It contains, (1) bedrock geology (12 rock types), (2) forest coverage (binary), (3) elevations, (4) aspect angles and (5) slope angles. We have the locations of 22 landslides occurred in 1964 and 51 landslides occurred in two time-periods, 1976 and 1996. Seven of the latter 51 landslides were occurred in the same areas of the 22 landslides occurred in 1964. The average size of the 73 landslides is approximately 15m x 15m covering 9 pixels. Among 2,400,000 pixels, 445, 164 pixels covered by the lake, the rivers were excluded in the study, and the remaining 1,954,836 pixels were included in the study. For each variable, the autocorrelation of two pixels depends on the distance between two pixels. Every variable has a strong spatial characteristic.

3. Favourability function model for Step 1

Consider that we have m map-layers containing the causal factors, which are known to correlate with the scarps of landslides in a study area \mathcal{A} . Consider a pixel i in \mathcal{A} with m pixel values, $X_1(i) = c_1, \dots, X_m(i) = c_m$, one for each map-layer, let $Y(i)$ represent the presence ($Y(i) = 1$) or absence ($Y(i) = 0$) of a past landslide and let $Z(i)$ represent the presence or absence of a future landslides at the pixel i .

As discussed in Chung (2002), suppose that, for every pixel $i \in \mathcal{A}$, we can construct a “favourability” function g :

$$g: (X_1(i), \dots, X_m(i)) \rightarrow R, \quad (3.1)$$

where $R = (-\infty, \infty)$ such that $g(c_1, \dots, c_m)$ represents (or measures) a relative level of hazard of the i^{th} pixel for given m pixel values, (c_1, \dots, c_m) for both past ($Y(i) = 1$) or future ($Z(i) = 1$) landslides. For the i^{th} pixel, we rewrite $g(c_1, \dots, c_m)$ as $g_i(Y=1 \text{ or } Z=1 | c_1, \dots, c_m)$ instead.

Then by computing $\hat{g}_i(Y=1 \text{ or } Z=1 | c_1, \dots, c_m)$ for every pixel in \mathcal{A} , we can construct a hazard map for \mathcal{A} .

4. Likelihood ratio function model for Step 1

Suppose that the study area is divided into two non-overlapping sub-areas, the scarps and the remaining areas. Suppose that the slope angles provide useful information to identify the scarps, and then the slope angle data of the scarps should have unique characteristics that are different from the data for the remaining areas. This suggests that the frequency distribution functions of the scarps and the remaining areas should be distinctly different as illustrated in Figure 1(a). The likelihood ratio function, which is the ratio of the two frequency distribution functions, can not only highlight this difference as illustrated in Figure 1(b) but also be the favourability function satisfying all three conditions discussed in the previous section.

To formalize the idea, let us consider a pixel p with m pixel values, c_1, \dots, c_m in the whole study area A consisting of two sub-areas, the scarps M and the remaining area \bar{M} .

$$\begin{aligned} M &: \text{set of pixels from the scarps,} \\ \bar{M} &: \text{set of pixels from the remaining area.} \end{aligned} \quad (4.1)$$

Let $f\{c_1, \dots, c_m | M\}$ and $f\{c_1, \dots, c_m | \bar{M}\}$ be the multivariate frequency distribution functions assuming that the pixel is from M , and from \bar{M} , respectively. Then the *likelihood ratio* (Kshirsagar, 1972; Cacoullos, 1973, McLachlan, 1992) at p is defined as: \bar{M}

$$\lambda_p(c_1, \dots, c_m) = \frac{f\{c_1, \dots, c_m | M\}}{f\{c_1, \dots, c_m | \bar{M}\}}. \quad (4.2)$$

For the slope angle whose distributions shown in Figure 1(a), the corresponding likelihood ratio function in logarithmic scale is illustrated in Figure 1(b). The ratio in Figure 1(b) obviously displays significant differences.

The discriminant analysis (Kshirsagar, 1972; Cacoullos, 1973; Chung, 1975; Chung, 1977) consists of estimating the likelihood ratio function $\lambda_p(c_1, \dots, c_m)$ in (2) based on the data from Table 1. To apply discriminant analysis, we assume that: (i) all m layers are based on continuous observations, (ii) $f\{c_1, \dots, c_m | M\}$ and $f\{c_1, \dots, c_m | \bar{M}\}$ are the normal density functions. Many statistical packages such as SPSS (SPSS, 1994) and S-Plus (Chambers and Hastie, 1992) provide solutions to the traditional discriminant analysis and the variation of the analysis. We take the estimates of $\lambda_p(c_1, \dots, c_m)$ in (4.2) as the favourability function \hat{g}_i ($Y=1$ or $Z=1 | c_1, \dots, c_m$) discussed in Section 3. We compute the estimates of $\lambda_p(c_1, \dots, c_m)$ for every pixel in the study area. The pixel with the largest estimate is considered as the most hazardous sub-area for future landslides according to this discriminant model.

When we consider several layers simultaneously in the study area, we now have m pixel values, c_1, \dots, c_m , at a pixel p . The likelihood ratio at p is the same as shown in (4.2). Suppose that the m layers provide "independent" sets of information over the scarps and the remaining area (i.e.,

we assume the conditional independence, as discussed Duda and et al. 1976; Heckerman, 1986; Spiegelhalter, 1986; Agterberg, et al. 1990; Chung and Fabbri, 1998), then (4.2) becomes,

$$\begin{aligned} \lambda_p(c_1, \dots, c_m) &= \lambda_p(c_1) \dots \lambda_p(c_m) \\ &= \frac{f\{c_1 | M\}}{f\{c_1 | \bar{M}\}} \dots \frac{f\{c_m | M\}}{f\{c_m | \bar{M}\}} \end{aligned} \quad (4.3)$$

The advantage of (4.3) over (4.2) is that it depends only on the univariate distribution function for each layer. The price of the advantage, however, is that the simplification requires the assumption of the conditional independence. Using (4.3), combining two different types (categorical and continuous) of data layers becomes a trivial matter.

To obtain the corresponding empirical distribution functions for $f\{c_i | M\}$ and $f\{c_i | \bar{M}\}$ from the data, we have employed the smoothed kernel method. The estimator of the likelihood ratio is obtained by:

$$\begin{aligned} \hat{\lambda}_p(c_1, \dots, c_m) &= \hat{\lambda}_p(c_1) \dots \hat{\lambda}_p(c_m) \\ &= \frac{\hat{f}\{c_1 | M\}}{\hat{f}\{c_1 | \bar{M}\}} \dots \frac{\hat{f}\{c_m | M\}}{\hat{f}\{c_m | \bar{M}\}}, \end{aligned} \quad (4.4)$$

where

$\hat{f}\{c_1 | M\}, \hat{f}\{c_1 | \bar{M}\}, \dots, \hat{f}\{c_m | M\}, \hat{f}\{c_m | \bar{M}\}$ are the corresponding empirical distribution functions. In this case, we take $\hat{\lambda}_p(c_1, \dots, c_m)$ in (4.4) as the favourability function \hat{g}_i ($Y=1$ or $Z=1 | c_1, \dots, c_m$). For every pixel, we compute $\hat{\lambda}_p(c_1, \dots, c_m)$. The pixel with the largest estimate is considered as the most hazardous sub-area for future landslides according to this model.

Using the 66 locations of the 73 landslides occurred during the past 38 years (1964 – 2002) five layers of geomorphological information related to the landslides in the area, we compute $\hat{\lambda}_p(c_1, \dots, c_m)$ in (4.4) for each of 1,954,836 pixels in the study area. According to the rank order of $\hat{\lambda}_p(c_1, \dots, c_m)$ for 1,954,836 pixels, we have divided the study area into 1000 classes. Each class contains 1,955 pixels (covers approximately 0.05 km²). 1,955 pixels with the highest $\hat{\lambda}_p(c_1, \dots, c_m)$ were assigned as the most hazardous predicted area in the study area. These classes are shown in Figure 2. As in the color legend consisting of 40 color bars in Figure 2, each color bar represents 25 classes of the 1000 original classes. The most hazardous 25 classes (48,871 pixels covers approximately 1.22 km² or 2.5% of the study area) were shown as purple and the subsequent most hazardous 48,871 pixels were shown as pink in Figure 2.

5. Estimation of conditional probability – Step 2

The first step is to construct a hazard map with a

number of hazard classes as similarly done by the geomorphologists or civil engineers. The second step is to estimate the probability in each class given a scenario or assumptions.

Let us take an example. Suppose that we build a house of size 10m x 25m (250 m²) within the most hazardous class (covers approximately 0.5 km²) of the 1000 classes in Figure 2. The next logical step is to estimate the conditional probability that the house will be affected by a future landslide within the next 35 years. We are proposing to estimate the probability empirically using the cross-validation technique.

Suppose that the time of landslide hazard study in La Baie is 1967 (35 years ago from today, 2002). In the study area, we know the locations of 22 landslides and we have five layers of geomorphological information. Using these 1967 data, we have computed $\hat{\lambda}_p(c_1, \dots, c_m)$ in (4.4) for each of 1,954,836 pixels in the study area. Similar to Figure 2, the 1,954,836 pixel values of $\hat{\lambda}_p(c_1, \dots, c_m)$ were sorted from the descending order, and then 1,954,836 pixels into 1000 hazard classes according to the descending order. For Figure 3, we have grouped 1000 classes into 40 groups of 25 classes each. In Figure 3, we have also shown 51 landslides occurred in 1976 and 1996 as black dots. The information on these 51 landslides was not used to construct Figure 3.

The first column in Table 1 represents the portion of the whole study area assigned as “hazard” area for future landslides. The first label “Top 1%” in the column is for the group of the most hazard 10 classes of the original 1000 classes and subsequent “1 – 2%” group is for the next 10 classes. To generate the second column in Table 1, in each of the 1000 classes, we have first made a cumulative count of the 51 landslides. For the classes without the landslides, instead of the cumulative counts, we have used interpolated values. Among the 1000 pairs, we have selected the 20 pairs shown in the second column of Table 1 and it constitutes 2/5th of red curve in Figure 4(a). It is termed as “prediction rate curve.”

To estimate the probability, we need more assumptions on the future landslides within the next 35 years. We need to have the “expected” number of future landslides in the area within the next 35 years and the “expected” size of the landslides. Since we had 51 landslides for the past 35 years in the study area and the average size of the past 51 landslides is approximately 15m x 15m, we will make the following additional assumptions:

- (i) 50 landslides will be occurred in the study area in the next 35 years;
- (ii) the average size of the 50 “future” landslides will be 15m x 15m.

From the assumptions in (5.1), the affected area by 50 landslides expected within the next 35 years is 50 x 15m x 15m or 450 pixels (size 5m x 5m). If we were to build a house of size 10m x 25m (250 m² or 10 pixels) in the most hazard 1% area (“Top 1% area), then the probability that the house will be a part of the whole affected area can be estimated by:

$$\text{An estimate} = 1 - (1 - \delta)^{10 (= \text{size of house})} \quad (5.2)$$

$$\text{where } \delta = \frac{450 (= \text{size of affected area})}{1954.35} \text{ probability ,}$$

“probability” equals to 0.28 shown in the corresponding row for “Top 1% area” of the second column in Table 1. The estimate is 6.26% shown in the corresponding row for “Top 1% area” of the third column. Similarly the numbers in the third column were generated from (5.2) using the corresponding probabilities in the second column in Table 1.

Theoretically speaking, the prediction rate curve must satisfy two conditions: (i) monotone increment function, and (ii) the increment rate (the target of the curve) must be monotone decrement function. Obviously, the red prediction rate curve in Figure 4(a) doesn’t satisfy the second condition. For that, we have fitted a linear exponential function for the red prediction rate curve in Figure 4(a) and it is shown in Figure 4(a) as a blue curve. The equation is:

$$\text{Fitted function} = 1 - e^{-0.17 - 7.15 \text{ area}} \quad (5.3)$$

The “area” in the equation represents the portion of the whole area as shown in the first column of the Table 1. The corresponding fitted values are shown in the 4th column in Table 1. Using the probabilities in the 4th column, the numbers in the 5th column were generated from (5.2).

The first 20 values in the 3rd and 5th columns were plotted and shown in Figure 4(b). Under the assumptions in (5.1), they are the estimated probabilities that a house of size 10m x 25m (250 m² or 10 pixels) in the corresponding 1% areas will be affected by future landslides within the next 35 years. Obviously while the 3rd column is based in empirical estimates, the 5th column is based on the fitted prediction rate curve shown as blue curve in Figure 4(a).

6. Acknowledgments

We wish to thank Dr. P. Didier, Geological Survey of Canada, who has provided the spatial data for the study. The study was also partly funded from a research grant provided to the Spatial Data Analysis Laboratory of Geological Survey of Canada by PCI Inc., Richmond Hill, Canada.

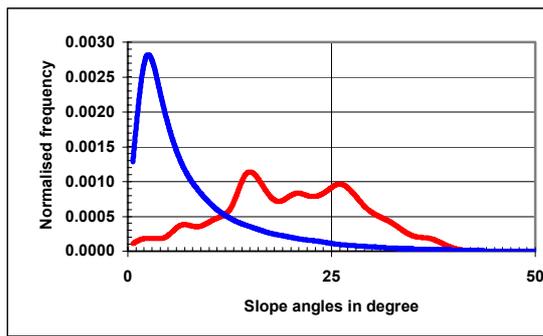
REFERENCES

- Agterberg, F.P., Bonham-Carter, G.F., and Wright, D.F., 1990. Statistical pattern integration for mineral exploration. In, Gaal, G. and Merriam, D.F., eds., *Computer Applications in Resource Estimation, Prediction and Assessment of Metals and Petroleum*. New York, Pergamon Press, p. 1-21.
- Cacoullos, T., 1973, *Discriminant Analysis and Applications*, Academic Press, New York, 434p.
- Chambers, J.M., and Hastie, T.J., 1992, *Statistical Models in S*, Wadsworth & Brooks/Cole, Pacific Grove, California, 608p.
- Chung, C.F., 1975, An application of classification analysis for Project Appalachia data. In *Proceedings of the 14th APCOM Symposium*, Pennsylvania State Univeristy, p.299-311.

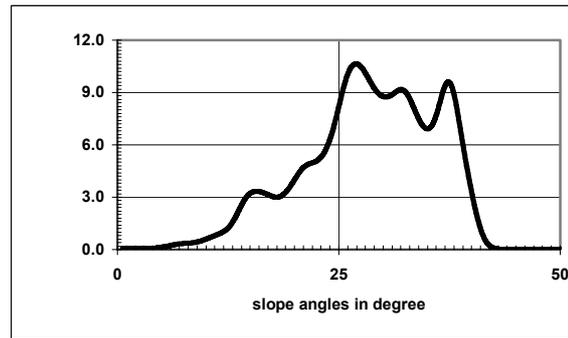
- Chung, C.F., 1977, An application of discriminant analysis for the evaluation of mineral potential. In Geological Survey of Canada paper 75-1C, Ottawa, Canada, p.141-148.
- Chung, C.F. and Fabbri, A.G., 1993, The representation of geoscience information for data integration. *Nonrenewable Resources*, v. 2, n. 2, p. 122-139.
- Chung, C.F., and Fabbri, A.G.: 1998, Three Bayesian prediction models for landslide hazard, In A. Bucciantti (ed.), Proceedings of International Association for Mathematical Geology 1998 Annual Meeting (IAMG'98), Ischia, Italy, pp. 204-211.
- Chung C.F. and Fabbri, A.G.: 1999, Probabilistic prediction models for landslide hazard mapping, Photogrammetric Engineering and Remote Sensing, 65-12, 1389-1399.
- Chung, C.F. and Fabbri, A.G.: 2001, Prediction model for landslide hazard using a Fuzzy set Approach, In M. Marchetti and V. Rivas (eds.), Geomorphology and Environmental Impact Assessment, Balkema, Rotterdam, in press.
- Chung, C.F. and Fabbri, A.G.: 2002, Validation for landslide hazard using a Fuzzy set Approach, In M. Marchetti and V. Rivas (eds.), Geomorphology and Environmental Impact Assessment, Balkema, Rotterdam, in press.
- Chung, C.F. and Perret, D.: 2002, Landslide hazard mapping in La Baie, Quebec, Canada, in preparation.
- Chung, C.F.: 2002, Two-step approach for spatial prediction models for landslide hazard mapping, in preparation.
- Duda, R.O., 1980, The Prospector systems for mineral exploration. SRI International Final report.
- Heckerman, D., 1986, Probabilistic interpretations for MYCIN's certainty factors. In L.N. Kanal and J.F. Lemmer Eds., *Uncertainty in Artificial Intelligence*. Elsevier Science Pub., North-Holland, pp. 167-196.
- Kshirsagar, A.M., 1972, *Multivariate Analysis*, Marcel Dekker Inc., New York, 534p.
- Panizza M., Corsini M., Soldati M. and Tosatti G.: 1998, Report on the use of new landslide susceptibility mapping techniques, In J. Corominas, J. Moya, A. Ledesma, J.A. Gili, A. Loret and J. Rius (eds.), New Technologies for Landslide Hazard Assessment and Management in Europe (NEWTECH). Final Report, October 1998 of CEC Environment Programme Contract ENV-CT96-0248, UPC, Barcelona, pp.13-31.
- Spiegelhalter, D.J., 1986, A statistical view of uncertainty in expert systems. In W.A. Gale, ed., *Artificial Intelligence and Statistics*, Addison-Wesley Pub., Reading, Mass., pp. 17-55.
- SPSS, 1994, *SPSS Professional Statistics 6.1*, SPSS Inc., Chicago, Illinois, 385p.
- Terlien, M.T.J., van Westen, C.J., and van Asch, T.W.J.: 1995, Deterministic modeling in GIS-based landslide hazard assessment, In A. Carrara and F. Guzzetti (eds.), Geographic Information Systems in Assessing Natural Hazards, Kluwer, Dordrecht, pp. 57-78.

Table 1. The first column represents the portion of the whole study area assigned as “hazard” area for future landslides. The first label “Top 1%” in the column is for the group of the most hazard 10 classes of the original 1000 classes and subsequent “1 – 2%” group is for the next 10 classes. As discussed in the text, the second column was generated by comparing the 1000 classes for Figure 3 and the 51 landslides occurred in 1976 and 1996. The 4th column was based a fitted function shown in (5.3) for the empirical values in the second column. The 3rd and 5th columns show, under the assumptions in (5.1), the estimated probabilities that a house of size 10m x 25m (250 m² or 10 pixels) in the corresponding 1% areas will be affected by a future landslides within the next 35 years using (5.2) and the probabilities shown in the 2nd and 4th, respectively. While the 3rd column is based in empirical estimates, the 5th is based on the fitted prediction rate curve shown as blue curve in Figure 4(a). The corresponding plots are shown in the Figure 4(b)

Portion of the study area assigned as hazard area	Cumulative portion of 51 landslides within the class	Empirical estimation based on the cumulative portion	Fitted function $1 - e^{-0.17 - 7.15 \text{ area}}$	Estimated from the fitted exponential function
Top 1%	0.2800	0.0626	0.2177	0.0362
1 – 2%	0.0373	0.0085	0.0540	0.0133
2 – 3%	0.0704	0.0161	0.0503	0.0124
3 – 4%	0.0476	0.0109	0.0468	0.0115
4 - 5%	0.0452	0.0104	0.0436	0.0107
5 – 6%	0.0480	0.0110	0.0406	0.0100
6 – 7%	0.0559	0.0128	0.0378	0.0093
7 – 8%	0.0186	0.0043	0.0352	0.0087
8 – 9%	0.0455	0.0104	0.0327	0.0081
9 – 10%	0.0120	0.0028	0.0305	0.0075
10 – 11%	0.0150	0.0034	0.0284	0.0070
11 – 12%	0.0207	0.0047	0.0264	0.0065
12 – 13%	0.0186	0.0043	0.0246	0.0061
13 – 14%	0.0186	0.0043	0.0229	0.0056
14 - 15%	0.0171	0.0039	0.0213	0.0053
16 – 17%	0.0100	0.0023	0.0198	0.0049
17 – 18%	0.0100	0.0023	0.0185	0.0046
18 – 19%	0.0090	0.0021	0.0172	0.0042
19 – 20%	0.0090	0.0021	0.0160	0.0040



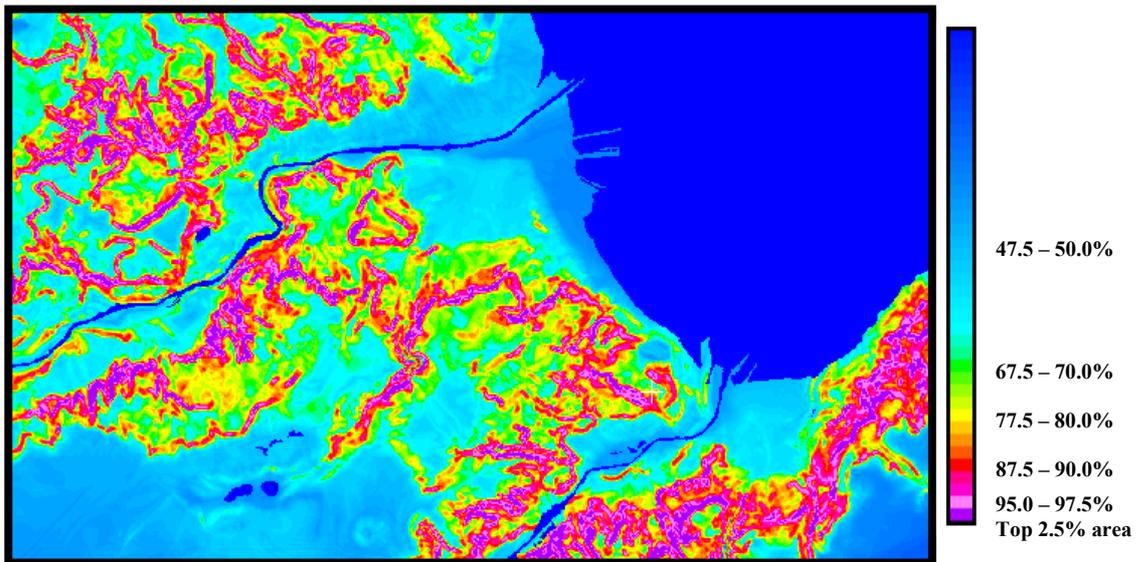
(a)



(b)

Figure 1. (a) Two empirical frequency distribution functions of 73 landslides area (in red) and the remaining area (in blue) using a kernel method. (b) The empirical likelihood ratio function based on two empirical distribution functions in (a).

5,356,213.5m N



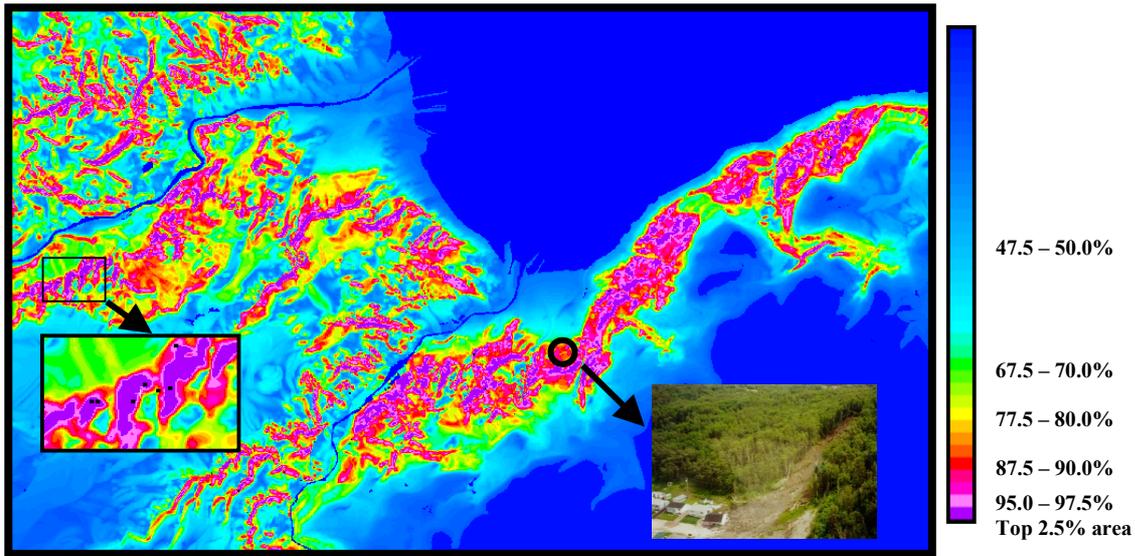
5,350,213.5m N

282,670.5m E

282,670.5mE

Figure 2. Landslide hazard prediction map based on 73 landslides (22 in 1967, 51 landslides in 1976 and 1996) and five layers (bedrock geology, forest coverage, elevation, aspect angle, slope angle maps) of geomorphological map information using likelihood ratio function model.

5,356,213.5m N

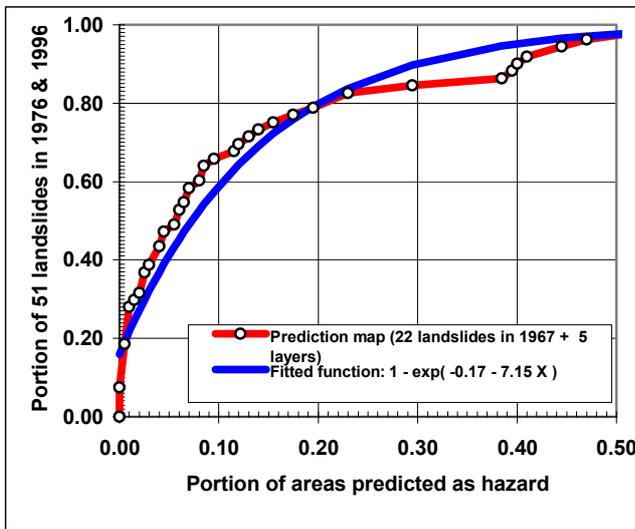


5,350,213.5m N

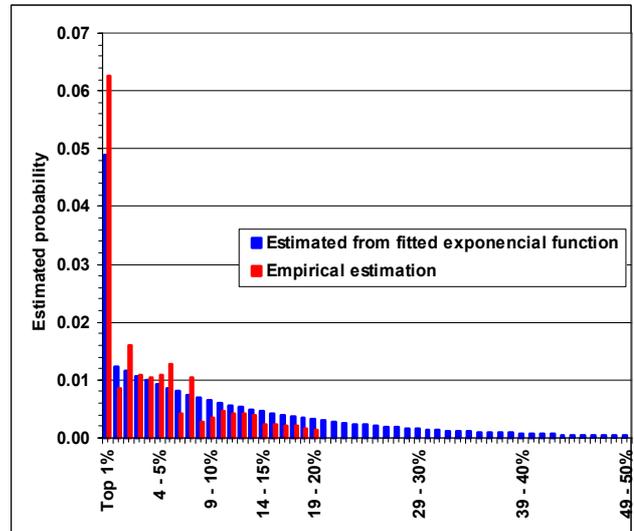
282,670.5m E

282,670.5m E

Figure 3. Landslide hazard prediction map based on 22 landslides occurred in 1967 and five layers (bedrock geology, forest coverage, elevation, aspect angle, slope angle maps) of geomorphological map information using likelihood ratio function model. The 51 black dots represent 51 landslides occurred in 1976 and 1996. The left side inset is an enlargement of a small area in black rectangle area in the middle left side. The right side inset with “Year 1996” is an image showing a photograph of a landslides occurred in 1996 at the black circle area in the middle area.



(a)



(b)

Figure 4. (a) Prediction rate curve for the prediction map shown in Figure 2. It was obtained by comparing the 1000 hazard classes generated for Figure 3 and the 51 landslides occurred in 1976 and 1996 as discussed in the text. The 20 pairs shown in the second column of Table 1 constitutes 2/5th of red curve. The fitted function shown in (5.3) is shown as blue curve. (b) It shows, under the assumptions in (5.1), the estimated probabilities that a house of size 10m x 25m (250 m² or 10 pixels) in the corresponding 1% areas will be affected by a future landslides within the next 35 years using (5.2) and the prediction rate curves shown in (a). Obviously while the red histogram is based in empirical estimates, the blue histogram is based on the fitted prediction rate curve shown as blue curve in Figure 4(a). The corresponding table values are shown in the 3rd and 5th columns in Table 1.