Jour. Korean Earth Science Society, v. 24, no. 5, p. 428–439, August 2003

Application of Spatial Data Integration Based on the Likelihood Ratio Function and Bayesian Rule for Landslide Hazard Mapping

No-Wook Park[1,*] · Kwang-Hoon Chi[1] · Chang-Jo F. Chung[2] · Byung-Doo Kwon[3]

[1]Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources, 30 Gajeong-dong, Yuseong-gu, Daejeon 305-350, Korea
[2]Geological Survey of Canada, 601 Booth Street, Ottawa, Canada K1A 0E8
[3]Department of Earth Science Education, Seoul National University, San 56-1, Shillim9-dong, Gwanak-gu, Seoul 151-748, Korea

우도비 함수와 베이지안 결합을 이용한 공간통합의 산사태 취약성 분석에의 적용

박노욱[1,*] · 지광훈[1] · Chang-Jo F. Chung[2] · 권병두[3]

[1]한국지질자원연구원 지질자원정보센터, 305-350 대전시 유성구 가정동 30번지
[2]Geological Survey of Canada, 601 Booth Street, Ottawa, Canada K1A 0E8
[3]서울대학교 지구과학교육과, 151-748 서울시 관악구 신림9동 산 56-1

Abstract: Landslides, as a geological hazard, have caused extensive damage to property and sometimes result in loss of life. Thus, it is necessary to assess vulnerable areas for future possible landslides in order to mitigate the damage they cause. For this purpose, spatial data integration has been developed and applied to landslide hazard mapping. Among various models, this paper investigates and discusses the effectiveness of the Bayesian spatial data integration approach to landslide hazard mapping. In this study, several data sets related to landslide occurrences in Jangheung, Korea were constructed using GIS and then digitally represented using the likelihood ratio function. By computing the likelihood ratio, we obtained quantitative relationships between input data and landslide occurrences. The likelihood ratio functions were combined using the Bayesian combination rule. In order for predicted results to provide meaningful interpretations with respect to future landslides, we carried out validation based on the spatial partitioning of the landslide distribution. As a result, the Bayesian approach based on a likelihood ratio function can effectively integrate various spatial data for landslide hazard mapping, and it is expected that some suggestions in this study will be helpful to further applications including integration and interpretation stages in order to obtain a decision-support layer.

Keywords: spatial integration, Bayesian, likelihood ratio, future landslide hazard

요 약: 여러 지질재해 중에서 산사태로부터의 피해를 최소화하기 위해서는 미래의 산사태에 대해 취약한 지역의 추정이 필요하다. 산사태 위험성의 정량적 분석을 목적으로, 본 논문에서는 확률론적 공간통합 방법인 베이지안 기법의 적용가능성에 대해서 논의하고자 한다. 우선 산사태 발생과 관련이 있는 여러 공간자료의 확률론적 표현을 위해 우도비 함수를 사용하였으며, 베이지안 결합 규칙을 이용하여 최종적으로 통합된 결과를 얻고자 하였다. 또한 미래의 산사태에 대한 의미있는 해석을 위하여 과거 산사태 공간 분포의 분할을 통한 검증을 수행하였다. 이러한 방법의 적용가능성을 검토하기 위하여 1998년 여름 산사태로 피해를 입은 경기도 장흥지역을 대상으로 사례연구를 수행하였다. 사례연구 수행 결과, 우도비에 기반한 베이지안 공간 통합 기법은 효율적으로 다양한 공간 자료를 통합할 수 있었으며, 검증결과는 해석과 의사결정 보조자료로 이용될 수 있을 것으로 기대된다.

주요어: 공간 통합, 베이지안, 우도비, 산사태 취약성

*Corresponding author: nwpark@kigam.re.kr
Tel: 82-42-868-3287
Fax: 82-42-867-0421

Introduction

Nowadays, the occurrences and extent of damage from geological hazards on human settlements are increasing. Even a small natural disaster can impact upon our human settlements very seriously and it will become progressively worse in the future. Landslides, as a geological hazard, are among the most costly catastrophic events in terms of human lives and infrastructure damage. Landslides may be induced by the progressive weakening of slope materials by slow natural processes, such as weathering and tectonic uplift, while others are induced by dynamic variables such as rainfall and local downpours (Zhou *et al.*, 2002). Especially, landslides triggered by heavy rainfall are the most common throughout Korea.

For the planning of future land use for economic activities and the prediction of possible landsliding zone, an essential component is the identification of those areas that are vulnerable to future possible landslides. Landslide occurrences are connected to a large number of geomorphological/environmental variables. Thus, we should consider multiple variables for landslide hazard mapping, and if we quantitatively connect these geomorphological characteristics with landslide occurrences, we can identify the area that is likely to be affected by future landslides.

For landslide hazard mapping, many researches have been carried out (Luzi and Florianna, 1996; Burton and Bathurst, 1998; Guzzetti *et al.*, 1999; Dai *et al.*, 2001). Traditional methods have used GIS to process the large bodies of data related to the landslide occurrences. Until now, traditional GIS functionality was based on the overlay analysis using the weights determined subjectively by experts. So it is severely affected by an erroneous input layer, ambiguous influence effects of data sets, inappropriate user-defined database query, and fuzziness of data sets themselves. In addition, most commercially available GISs do not provide information integration and are developed with insufficient mathematical understanding of the data. Thus, insufficient considerations for geoscience data sets may result in severely erroneous decision-making. Hence, we need more systematic usage of spatial data and methodologies that quantify the spatial relationships and integrate them efficiently in order to obtain the most reasonable interpretation.

Until now, most researches have focused on constructing quantitative models or methods and generation of maps showing landslide hazard. However, all prediction results related to future events are always subject to uncertainties. Thus, the predictions not only identify vulnerable areas, but also estimate the uncertainties associated with the prediction (Chung and Fabbri, 2003). However, proper interpretation and quantitative evaluation of prediction have not been fully considered in landslide hazard mapping.

In this paper, we introduce a probabilistic approach which implicitly assumes that most of the information on which decision-making is based is probabilistic in nature and that precise probability judgments can be formulated for each problem's hypothesis. In particular, we employed a Bayesian approach for landslide hazard mapping that can connect the quantitative relationship among multiple spatial data sets related to landslide occurrences and the degrees of uncertainty associated with the data.

In previous work on probabilistic methods for landslide hazard mapping, Chung and Fabbri (1999) constructed a probabilistic model based on joint conditional probability and showed that this model can be effectively used for landslide hazard mapping. In generic mathematical/statistical models for landslide hazard mapping, it is assumed that there are distinctive differences between landslide areas and non-landslide areas. The joint conditional probability only considers the portion of landslides in a certain attribute at once. So it cannot consider the effects of the remaining non-landslide areas. In addition, our main objective for landslide hazard assessment is to estimate relative hazard level within the study area. That is, we wish to separate the hazardous sub-areas affected by landslides and the non-hazardous sub-areas not affected by landslides. In such cases, the ratio of proportions is a useful descriptive measure. By adopting this idea, in this paper, we address the likelihood ratio function that can highlight these differences. Finally, as an essential part of landslide hazard mapping, we carry out a validation approach in order to evaluate the significance of the prediction results.

This paper is structured as follows. In the next two sections, we present the general concepts of spatial data integration, the Bayesian approach, and a likelihood ratio function. Then, a case study from Jangheung, Korea is described to illustrate the schemes proposed here. Specifically, in addition to generating a landslide hazard map, validation of a hazard map is emphasized for quantitative assessment of the uncertainties related to prediction. Finally, we conclude with discussion and remarks.

Bayesian Spatial Data Integration

Spatial Data Integration

Spatial data integration covers a very wide domain and so the term "data integration" has various definitions and terms according to the field in which it is applied. Spatial data integration is the formal framework that expresses the means and tools for the alliance of data originating from different sources (Wald, 1999). In particular, in geological applications (e.g. landslide hazard mapping or mineral potential mapping), spatial data processing involves several steps; data acquisition, data pre-processing, information representation, integration or fusion, visualization of the integrated information with respect to the target proposition, interpretation, and decision making (Fig. 1). Among these different steps, in a mathematical form, information representation and integration may also be explained as a mapping or a transformation between the raw spatial data and the integrated information with respect to a chosen target proposition. For information representation, the sureness that the target proposition is true is expressed in terms of probability, belief, and possibility, using the quantitative relationships between input spatial data and the known occurrences. Then the integration of representations into one single function can be done by various integration methods, depending on their mathematical frameworks (Moon, 1993). Among various data integration methods, we will focus on the Bayesian probabilistic approach for data representation and integration of data. In the next section, a general

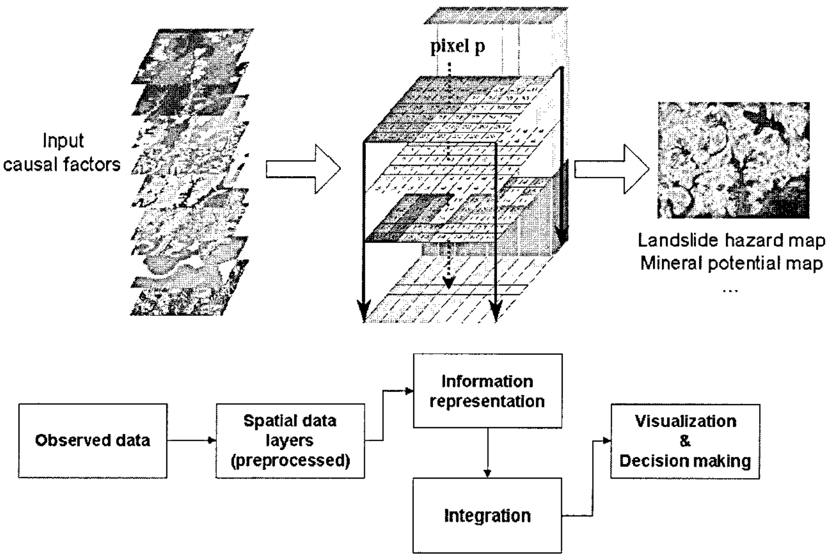


Fig. 1. Schematic flowchart for spatial data integration.

rationale of Bayesian data integration will be given.

Bayesian Approach

Bayesian approach provides a formalism for reasoning about partial beliefs under conditions of uncertainty. In this approach, propositions are given numerical parameters signifying the degree of belief accorded them under some bodies of knowledges, and the parameters are combined and manipulated according to the rules of probability theory (Sivia, 1996; Pearl, 1997).

In the Bayesian approach, sureness measures obey the basic axioms of probability theory and the basic expressions are statements about the prior probability, likelihood, and posterior probability. One of main concepts for applications is that the prior probability is successively updated with the addition of new evidence, so that the posterior probability from adding one piece of evidence can be treated as the prior probability for adding a new piece of evidence (Bonham-Carter, 1994; Kwon and Oh, 2002).

In our approach for landslide hazard mapping, a Bayesian model is based on the conceptual idea of expressing the landslide hazard in terms of probability with respect to spatial data and of combining them by the Bayesian combination rule. Suppose that m spatial data related to landslide occurrences are assembled for the identification of the vulnerable areas for a specific future landslide type in a study area A. Each layer of spatial data is regarded as a piece of evidence Ei $(i = 1, 2, \Lambda, m)$ for the target proposition such as "At each pixel p, it will be affected by future flow type landslides", denoted by T_P. The sureness that the target proposition is true is expressed in terms of conditional probability in a Bayesian probabilistic framework. In this interpretation, what we want to have is the joint conditional probability, denoted by Prob $\{T_P \mid E_1, E_2, \Lambda, E_m\}$ $(i = 1, 2, \Lambda, m)$. This joint conditional probability value indicates how each of the m spatial data, E_i, supports the sureness that the proposition is true. Using Bayes's theorem, we can obtain the joint conditional probability as follows:

$$\mathrm{Prob}\{T_p|E_1,\ E_2,\ \Lambda,\ E_m\} = \frac{\mathrm{Prob}\{T_p\}\mathrm{Prob}\{E_1,\ E_2,\ \Lambda,\ E_m|T_p\}}{\mathrm{Prob}\{E_1,\ E_2,\ \Lambda,\ E_m\}} \quad (1)$$

where, $\mathrm{Prob}\{T_P\}$ is the prior probability that each pixel p in the study area A will be affected by future landslides before we have any spatial data E_i $(i = 1, 2, \Lambda, m)$. It is a constant for all pixels, since it relates to neither any specific factor nor any specific pixels. $\mathrm{Prob}\{E_1, E_2, \Lambda, E_m\}$ is the probability that each pixel p in A has the spatial data. $\mathrm{Prob}\{E_1, E_2, \Lambda, E_m|T_P\}$ is the likelihood that the spatial data will materialize if T_P is true. $\mathrm{Prob}\{T_P|E_1, E_2, \Lambda, E_m\}$ is the posterior probability.

In this paper, instead of using the conditional probability directly, we used the likelihood ratio function. Unlike a tradition conditional probability approach, the likelihood ratio function has the advantage of considering the relative risk.

The basic concept of the likelihood ratio function is as follows. To identify the hazard areas for future landslides, it is necessary to separate the hazardous sub-areas affected by landslides and the non-hazardous sub-areas not affected by landslides. Suppose that the study area is divided into two non-overlapping sub-areas. If spatial data are to provide useful information, then the data from the hazardous sub-areas should have unique characteristics that are different from the data from the non-hazardous sub-areas. This suggests that the frequency distributions of the hazardous and the non-hazardous sub-areas should be distinctly different. The likelihood ratio function, which is the ratio of the two frequency distributions, can highlight this difference (Fig. 2). Details of the conceptual background and applications are discussed in some references (Duda *et al.*, 1976; Chung and Fabbri, 1998).

The joint likelihood ratio at p, λ is defined as:

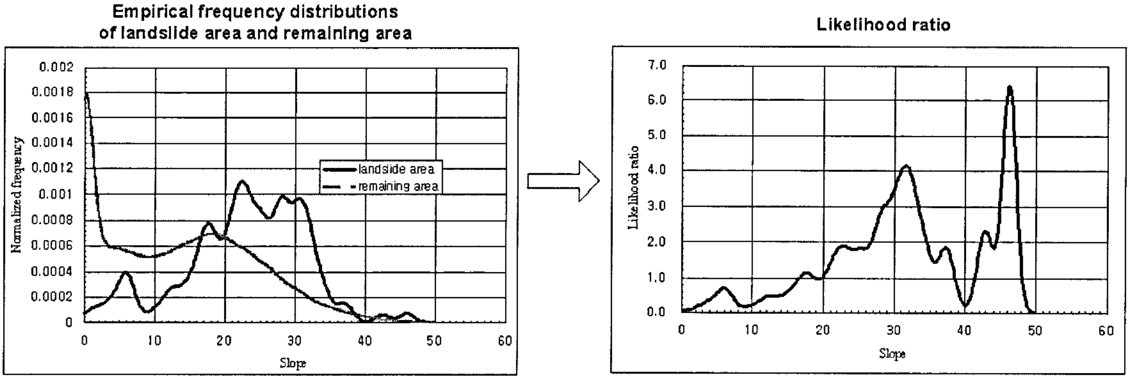


Fig. 2. Basic concept of the likelihood ratio function.

$$\lambda = \frac{\mathrm{Prob}\{E_1,\ E_2,\ \Lambda,\ E_m|T_p\}}{\mathrm{Prob}\{E_1,\ E_2,\ \Lambda,\ E_m|\overline{T}_p\}} = \frac{1-\mathrm{Prob}\{T_p\}}{\mathrm{Prob}\{T_p\}} \cdot \frac{\mathrm{Prob}\{T_p|E_1,\ E_2,\ \Lambda,\ E_m\}}{1-\mathrm{Prob}\{T_p|E_1,\ E_2,\ \Lambda,\ E_m\}} \tag{2}$$

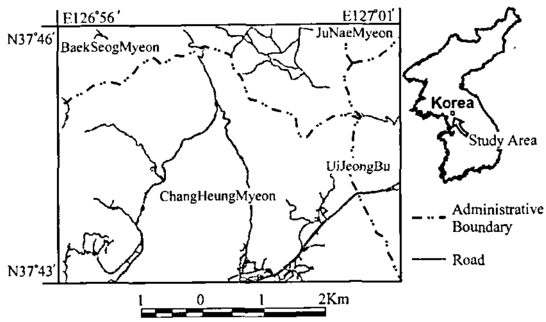
where $\overline{T}_p$ denotes the proposition that at each pixel p it will not be affected by future flow-type landslides.

The joint likelihood ratio is the ratio function of the joint conditional probability and prior probability discussed above. It highlights the relative difference between the joint conditional probability and the prior probability at p, even more than the simple ratio itself, by multiplying the ratio of the remaining probabilities (Chung and Fabbri, 1998). It is always positive and ranges from zero to infinity. It is greater than one if the relations between T_p and $\overline{T}_p$ are positively associated, it is one if they are independent, and less than one if they are negatively associated. The more the likelihood ratio exceeds one, the stronger the relationship between two patterns will be.

To calculate the joint likelihood ratio, we need to obtain the joint conditional probability in advance. In practice, however, the joint conditional distribution is rarely specified explicitly. So, as an approximation, it is assumed that spatial data provide independent sets of information, namely the conditional independence assumption. Under the conditional independence assumption, we can simplify the mathematical analysis and computations, and the joint likelihood ratio can be expressed as a product of factor-specific conditional distributions, i.e., the bivariate conditional probability at each layer. In an evidence sense, the independence assumption is related to the conceptual understanding of how spatial data are related to landslide occurrences rather than to actual spatial patterns in spatial data. The validity of an independence assumption is difficult to verify. And this assumption is often the only alternative because the form of the multivariate distribution is unknown.

$$\lambda = \frac{\mathrm{Prob}\{E_1,\ E_2,\ \Lambda,\ E_m|T_p\}}{\mathrm{Prob}\{E_1,\ E_2,\ \Lambda,\ E_m|\overline{T}_p\}} = \frac{\mathrm{Prob}\{E_1|T_p\}}{\mathrm{Prob}\{E_1|\overline{T}_p\}} \cdot \frac{\mathrm{Prob}\{E_2|T_p\}}{\mathrm{Prob}\{E_2|\overline{T}_p\}} \Lambda \frac{\mathrm{Prob}\{E_m|T_p\}}{\mathrm{Prob}\{E_m|\overline{T}_p\}} = \prod_{i=1}^{k} \lambda_i \tag{3}$$

where, $\lambda_i = \frac{\mathrm{Prob}\{E_i|T_p\}}{\mathrm{Prob}\{E_i|\overline{T}_p\}} = \frac{1-\mathrm{Prob}\{T_p\}}{\mathrm{Prob}\{T_p\}} \cdot \frac{\mathrm{Prob}\{T_p|E_i\}}{1-\mathrm{Prob}\{T_p|E_i\}}$



(a)

(b)

Fig. 3. (a) Location map of the study area, (b) landslide scars that occurred in the study area.

Case study

The Jangheung area in Korea, which had considerable landslide damage following heavy rain in 1998, was selected as the study area (Fig. 3). Intense rainfall between August 4 and 9, 1998 had triggered many landslides in the study area. In the study area, the landslides were mainly flows that occurred during 3-4 hours of high intensity rainfall, or shortly after.

A flow diagram of the study is shown in Fig. 4. For landslide-hazard mapping, first, spatial data related to landslide occurrences were constructed. Then, the landslide hazard was analyzed using a Bayesian method based on likelihood ratio. Finally, through a validation procedure, we estimated the uncertainties in terms of the probabilities of the occurrences.

Input Spatial Data

First, an inventory of landslides was examined using change detection analysis based on high-resolution panchromatic satellite remotely sensed images such as two IRS 1-C images and KOMPSAT (KOrea Multi-Purpose SATellite) EOC image, acquired on 5 June 1998, 12 October 1998, and 2 February 1999, respectively. Before comparison, an image contrast enhancement technique was applied to each image. After applying a spectral normalizing algorithm for reducing the spectral discrepancy caused by differences in acquisition dates, the image difference technique was applied to obtain a map showing changed areas and non-changed areas. In Chi *et al.* (2001), a total of 359 landslides had been detected. However, the landslide locations detected from remotely sensed images may include other land-covers such as roads and tombs. So the locations were then verified by fieldwork, and finally a total of 332 landsides were mapped and the topographically highest 10% of the scars of the landslides were considered as trigger areas (Fig. 5 (a)).

As the basic analysis tool, GIS was used for spatial data management and manipulation. First, a raster-based GIS spatial database including spatial data that are relevant to landslide occurrences was constructed. Since the study area mainly consists of gneiss in lithology and most landslides occurred in that area, in the whole database only gneiss areas were considered.

The study area covers approximately 37.29km^2 and consists of 1,491,443 pixels, with a pixel size of 5m by 5m. The spatial database consists of 6 layers. They are (1) the spatial distribution of the 332 landslides; (2) slope; (3) aspect; (4) forest type; (5) soil; and (6) distances from drainage patterns (Fig. 5). The slope and aspect were obtained from the Digital Elevation Model (DEM) of the study area. Forest type information was extracted from a 1 : 25,000 scale forest map. Soil information was extracted from a 1 : 50,000 scale reconnaissance soil map. In addition, distances from drainage patterns extracted from a digital topographic map were used in the analysis.

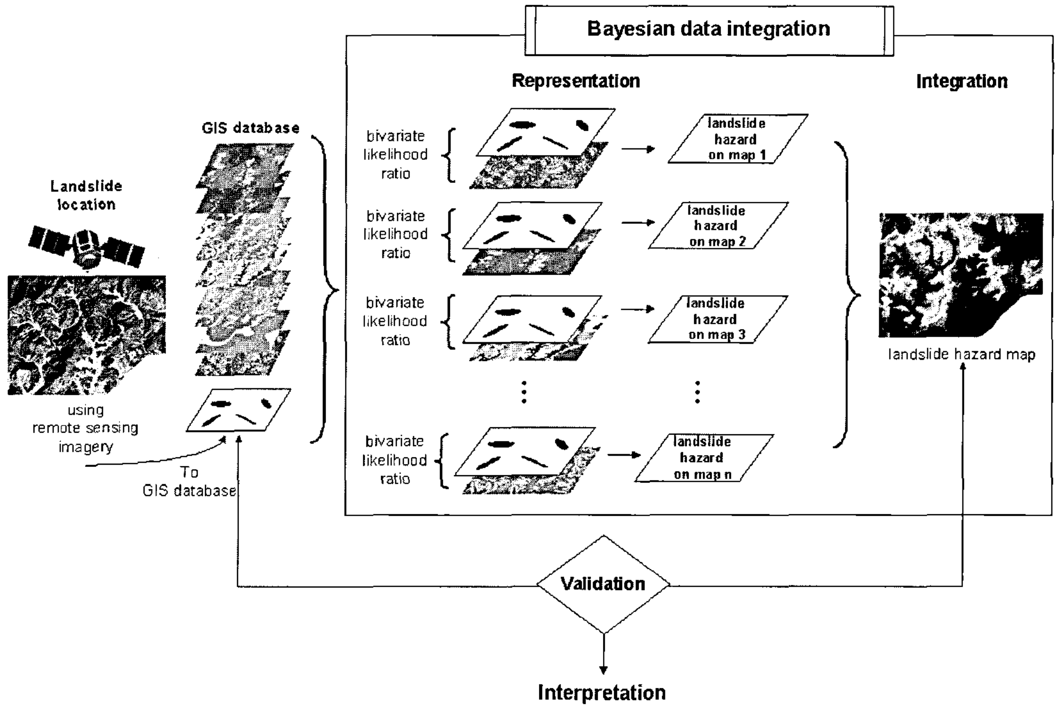


Fig. 4. Schematic diagram showing the processing flow used in this study.

Information representation

We estimated the likelihood ratio functions for each of spatial data using a quantitative relationship between past landslides and spatial data. A priori probability $\mathrm{Prob}\{T_p\}$ and conditional probability $\mathrm{Prob}\{E|T_p\}$ were calculated by using the following two formulas:

$$\mathrm{Prob}\{T_p\} = \frac{\text{the number of pixels in all past landslides}}{\text{the number of pixels in the whole study area}} \tag{4}$$

$$\mathrm{Prob}\{E_i|T_p\} = \frac{\text{the number of pixels in past landslides that occurred in certain class attribute } E_i}{\text{the number of pixels in certain class attribute } E_i} \tag{5}$$

The likelihood ratio functions for spatial data are shown in Table 1. As discussed in Section 2 above, if the value is greater than one, it means a higher association, and a value less than one means a lower association. In the slope map, the likelihood ratio increased according to the slope angle. In particular, classes whose slope angle was higher than 20 degrees showed likelihood ratio values greater than one. This means that most landslides had occurred in areas where the slope angle is greater than 20 degrees. In the aspect map, the likelihood ratio values on east-, southeast-, south- and south-west-facing hill slopes were high. These results may be caused by the difference in hours of sunshine according to the aspect. In the forest type map, the likelihood ratio was the highest for Korean nut pine. This result corresponds with other reported results that landslide occurrence probability is higher in a coniferous forest than in a broadleaf forest (Lee and Min, 2001). On the soil map, the likelihood ratio values in the soil IV and V rate classes were higher than those of another classes. According to the increase in rate, the soil has little soil moisture and has a low thickness. This is related to an

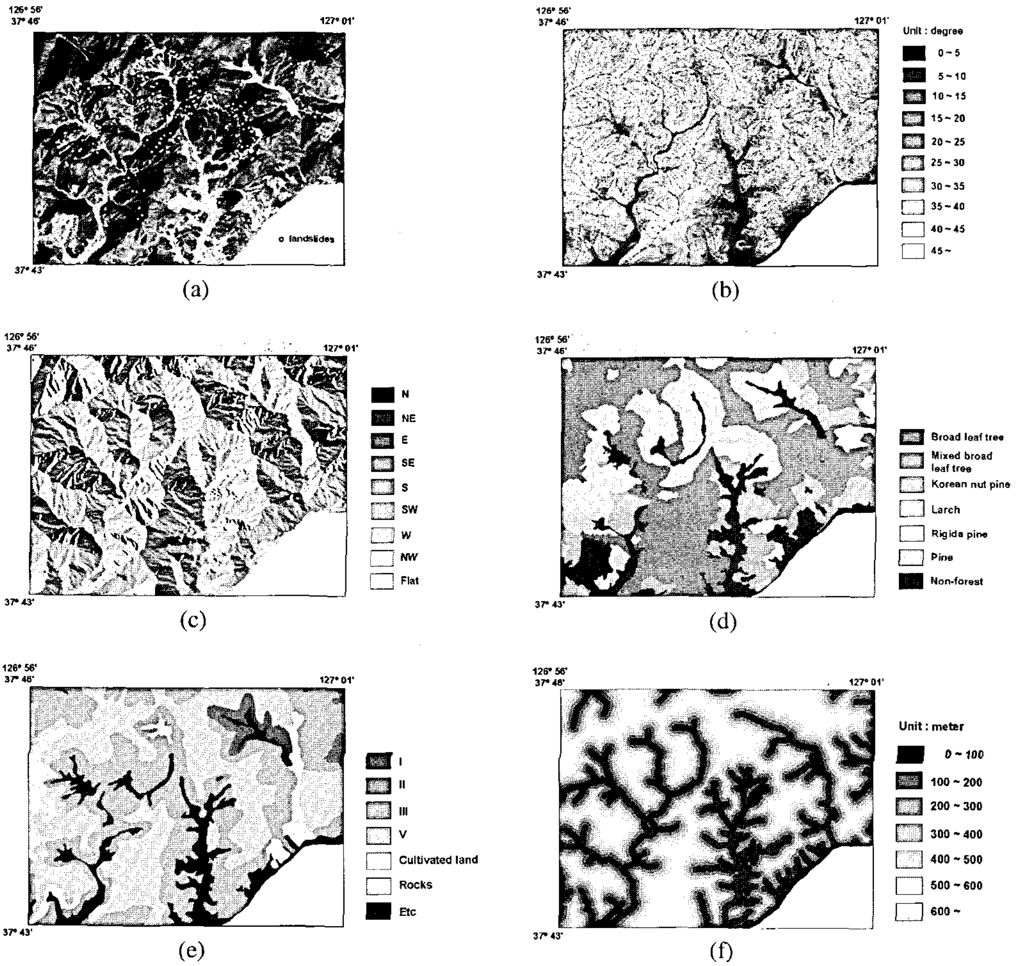


Fig. 5. Input spatial data, (a) landslide locations draped over the KOMPSAT EOC imagery, (b) slope map, (c) aspect map, (d) forest type map, (e) soil map, (f) distances from drainage pattern.

increase of unit weight and of shear stress of soil due to pore-water increase. For an increase in distance from a drainage pattern, the likelihood ratio value increased.

Integration

After obtaining the likelihood ratio function for each of the spatial data, the joint likelihood ratio functions were calculated using a Bayesian combination rule.

We used rank order statistics to visualize the computed joint likelihood ratio functions and to express landslide hazard in terms of relative values in the study area. We first computed the score for each pixel and then sorted all scores in increasing order to determine the ranks of the scores. The pixel that has the smallest score (the smallest prediction value) has rank one, and the pixel that has the largest score has the maximum rank. Then the ranks are normalized so that the maximum value is 1 or 100%. The pixel with an index of 100% has the largest score of the integrated result. If the pixels have indexes of 99.5%, it means that the ranks of their function scores are within the top 0.5% (99.5-100%) in the study area. Because landslide hazard has a ranking equal sub-area in the landslide hazard map, it is the same as the percentage of the study area used (i.e. equal-area classes).

Table 1. Likelihood ratio value of each layer

Layer	Class	Value	Layer	Class	Value
Slope (unit: degree)	0-5	0.055	Forest type		
	5-10	0.057			
	10-15	0.190		Broad leaf tree	0.788
	15-20	0.791		Mixed broad leaf tree	0.464
	20-25	1.218		Korean nut pine	2.540
	25-30	1.051		Larch	0.113
	30-35	1.656		Rigida pine	0.717
	35-40	2.400		Pine	0.001
	40-45	2.174		Non-forest	0.070
	45-	2.337			
Aspect	N	0.626	Distances from drainage pattern (unit: meter)		
	NE	1.013		0-100	0.221
	E	1.455		100-200	0.869
	SE	1.421		200-300	1.340
	S	1.165		300-400	1.404
	SW	1.255		400-500	1.536
	W	0.657		500-600	1.923
	NW	0.529		600-	1.076
	Flat	0.001			
Soil	II	0.001			
	III	0.755			
	IV	1.693			
	V	1.679			
	Rocks	0.167			
	Cultivated land	0.001			
	Etc	0.051			

These indexes over the study area constitute the landslide hazard map.

Through this process, the final landslide hazard map was generated, as shown in Fig. 6. Especially, high hazard areas are located in the north-eastern part of study area. This part mainly consists of steeply sloping, Korean nut pine, and IV or V rate soil class whose likelihood ratio values are high.

Validation

The final goal of landslide hazard mapping is to identify the area that is likely to be affected by future landslides. Therefore, landslide hazard mapping is a kind of a predictor for an unknown future event. So to generate a significant landslide hazard map, we should show how successful the prediction would be with respect to future landslides. The traditional validation approach is to compare the integrated results with the occurrence of the past landslides. However, a conceptually proper approach should compare the integrated results with the occurrences of future landslides. Since we do not have any information about what will happen in the future that is relevant to time and/or space, we need to use some part of the past landslides as if they represent future landslides.

Chung and Fabbri (2003) discussed the problem of providing measures of significance of prediction results when hazard maps were generated. For this, they carried out cross-validation, an empirical approach that tests the prediction map experimentally. In their approach, they used some part of the past landslides as if they represent future landslides and then one subset was used for obtaining a hazard map; the other subset was compared with the hazard map for validation. For partitioning the past landslides, they proposed three partitioning techniques: time, space, random partitioning. In time

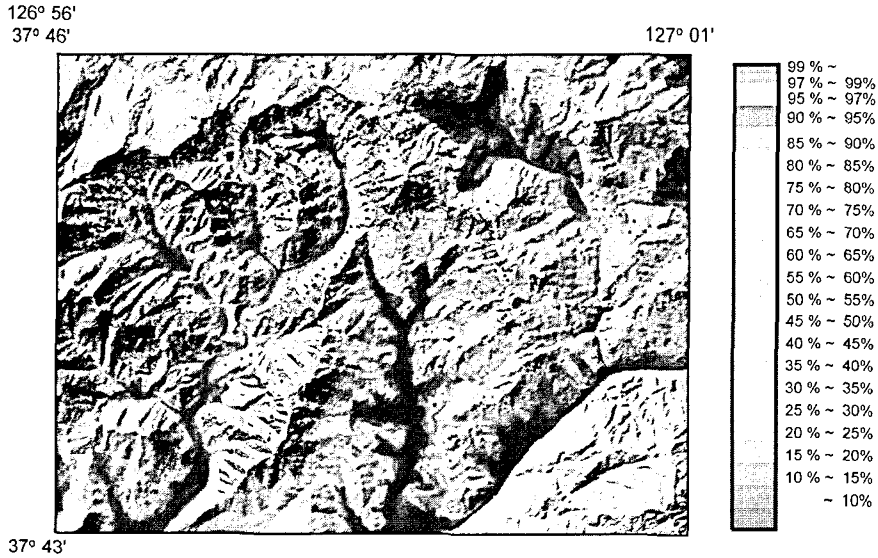


Fig. 6. Landslide hazard map based on likelihood ratio function. The background is a shaded relief image and black dots indicate the past landslides.

partitioning approach, we can estimate the probability of the occurrences of future landslides within a certain time constrain. In space partitioning approach, we can extend the current prediction model in the study area to the neighboring areas or similar environment areas. In random partitioning approach, the past landslides are randomly divided into two groups. In the study area, the landslides were induced by one time event, a heavy rainfall during some period in 1998 and we have no records of landslides which had occurred previous or after 1998. We also could not get any information on neighboring areas at the time of writing this paper. So it is not feasible to carry out time or space partitioning approach. Instead, we carried out a random partitioning approach for validation.

In this study, to evaluate the landslide hazard map in Fig. 6, we randomly divided the past 332 landslides into two groups. One subset (166 landslides, estimation set) was used as the estimation data set to construct probabilistic relationships between the landslides and the input data set, and used to generate the landslide hazard map. The integrated landslide hazard map based on those relationships was then evaluated by comparing the map pattern of the integrated hazard classes with the distribution of the other subset (166 landslides, validation set), assuming that the landslides have not yet occurred.

To quantitatively find how good the prediction is, we computed the prediction rate curve proposed by Chung and Fabbri (1999). Prediction rate is the measurement of how well the model predicts the distribution of future landslides. To calculate the prediction rate, we first counted the number of pixels of validation landslides in the landslide hazard level whose value is larger than (100 minus a certain value) %. Then the number was divided by the total pixel numbers of validation landslides in order to obtain a normalized prediction rate. The prediction rate curve is the cumulative version of the prediction rate. It has the form y = function (x) (Fig. 7). Here, x, ranging from 0 to 100%, is the percentage of relative landslide hazard and corresponds to the legend in the prediction map, as shown in Fig. 6. And y is the percentage of occurrences predicted within the most favorable x of the study area.

To compare the prediction power of the likeli-

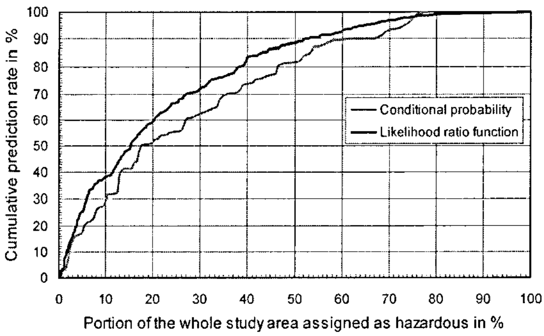


Fig. 7. Prediction rate curve for future landslide hazard.

hood ratio function with that of the conditional probability, two prediction rate curves were computed (Fig. 7). The likelihood ratio function showed higher predictive capability than the conditional probability. For instance, at the most hazardous top 10% areas, the prediction powers of the likelihood ratio function and the conditional probability function were about 39% and 30%, respectively. For the future landslides, if we assume that the type and size of the future landslides are identical to the past landslides, and take the most hazardous 10% of the area of the corresponding integrated image, then we may estimate that 39% of the future landslides will be located in the prediction result obtained from the likelihood ratio function. Since we use the same spatial data integration method and the same data in the study area, the prediction rate curve can be used to interpret the uncertainty in the integrated map generated by using all past landslides. How to prepare a landslide hazard map is important, but it is not a final step as far as the interpretation is concerned. Through the above validation procedure, we can not only evaluate a landslide hazard map quantitatively, but also get a meaningful interpretation with respect to the future landslide occurrences from the hazard map.

Conclusions

To identify areas that are susceptible to landslides, we should not only consider quantitative relationships between spatial data that represent the physical conditions of landslide and landslide occurrences, but also combine them effectively.

In this study, we applied a Bayesian data integration approach to landslide hazard mapping using multiple spatial data sets, and outlined the areas that will be affected by landslides. This approach includes mathematically proper representation of the information from different data sets and an effective framework for efficient combination of the evidence from each data set to obtain a reasonable and realistic interpretation. By computing the likelihood ratio functions of input data sets, we could obtain quantitative relationships between input spatial data and past landslides. To assess quantitatively the prediction powers of the prediction map, a cross-validation approach was also performed, with which we can evaluate the prediction results quantitatively. This approach can be considered as a decision-support layer from the viewpoint of GIS, with additional quantitative evidence not represented in the integrated landslide hazard map. In addition, this validation procedure can be used for quantitative comparison of various spatial data integration methods by computing the prediction rates. The likelihood ratio function showed the better prediction power that the results using the traditional conditional probability in this study. However, this result does not indicate that the likelihood ratio function shows better predictive capability than the conditional probability in all cases. So more research should be devoted to extensive experiments in several study areas to strengthen the situation here identified.

In this study, we only investigated the probabilistic integration method for landslide hazard mapping. For the future works, we will focus on constructing other spatial data integration methods, e.g. fuzzy logic and Shafer's theory of evidence. Especially, the validation procedure will play a crucial role in evaluating various methods.

Acknowledgment

This study was supported by the Korean Ministry of Science and Technology. The digital data sets in the study area were kindly provided by Mr. Jin-Soo Shin. The authors thank three reviewers, Professor Kiwon Lee of Hansung University, Dr. Seok-Hoon Oh of METRI/KMA, and one anonymous reviewer for their comments and suggestions that have improved an earlier version of this manuscript.

References

Bonham-Carter, G.F., 1994, Geographic information systems for geoscientists: modeling with GIS. Pergamon, New York, 398 p.

Burton, A. and Bathurst, J.C., 1998, Physically based modeling of shallow landslide sediment yield at a catchment scale. Environmental Geology, 35, 89-99.

Chi, K.-H., Shin, J.-S., and Park, N.-W., 2001, Quantitative evaluation of GIS-based landslide prediction model using prediction rate curve. Korean Journal of Remote Sensing, 17, 199-210.

Chung, C.F. and Fabbri, A. G., 1998, Three Bayesian prediction models for landslide hazard. Proceedings of International Association for Mathematical Geology Annual Meeting (IAMG'98), Ischia, Italy, 204-211.

Chung, C.F. and Fabbri, A.G., 1999, Probabilistic prediction models for landslide hazard mapping. Photogrammetric Engineering and Remote Sensing, 65, 1389-1399.

Chung, C.F. and Fabbri, A.G., 2003, Validation of spatial prediction models for landslide hazard mapping. Natural Hazard (in press).

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

Duda, R.O., Hart, P., and Nilsson, N., 1976, Subjective Bayesian methods for rule-based inference systems. Proceedings of the 1976 National Computer Conference, 45, AFIPS, 1075-1082.

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

Kwon, B.-D. and Oh, S.-H., 2002, Bayesian inversion of gravity and resistivity data: detection of lava tunnel. Journal of the Korean Earth Science Society, 23, 15-29.

Lee, S. and Min, K., 2001, Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental Geology, 40, 1095-113.

Luzi, L. and Florianna, P., 1996, Application of statistical and GIS techniques to slope instability zonation (1 : 50,000 Fabriano geological map sheet). Soil Dynamics and Earthquake Engineering, 15, 83-94.

Moon, W., 1993, On mathematical representation and integration of multiple spatial geoscience data sets. Canadian Journal of Remote Sensing, 19, 251-255.

Pearl, J., 1997, Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers, San Mateo, Calif., 552 p.

Sivia, D.S., 1996, Data Analysis: A Bayesian Tutorial. Oxford University Press, New York, 240 p.

Wald, L., 1999, Some terms of reference in data fusion. IEEE Transactions on Geoscience and Remote Sensing, 37, 1190 -1193.

Zhou, C.H., Lee, C.F., Li, J., and Xu, Z.W., 2002, On the spatial relationship between landslides and causative factors on Lantau Island, Hong Kong. Geomorphology, 43, 197-207.

2003년 5월 15일 원고 접수
2003년 6월 30일 수정원고 접수
2003년 7월 25일 원고 채택