# FUZZY LOGIC INTEGRATION FOR LANDSLIDE HAZARD MAPPING USING SPATIAL DATA FROM BOEUN, KOREA

Kwang-Hoon Chi [a], No-Wook Park [a,]*, Chang-Jo Chung [b]

a National Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources, Daejon 305-350, Korea - nwpark@kigam.re.kr
b Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario K1A 0E8, Canada - chung@gsc.nrcan.gc.ca

**KEY WORDS:** Fuzzy logic, Data Integration, Validation, Future Landslides

**ABSTRACT:**

Since the early 1990s, the mathematical models have been developed and applied to landslide hazard mapping using GIS. Among various models, this paper discusses the effectiveness of fuzzy set theory for landslide hazard mapping. In this study, we collected several data sets related to landslide occurrences in Boeun, Korea, and then digitally represented as the fuzzy membership functions. To integrate them, fuzzy inference networks by using a variety of different fuzzy operators, especially combination of fuzzy OR and fuzzy $\gamma$ operator, are designed, and experiments are carried out. Owing to the cross-validation based on the spatial portioning of the landslide distribution, we could quantitatively compare with various fuzzy inference networks designed for the influence of choice of $\gamma$ value. The results show that the fuzzy set theory can integrate effectively various spatial data for landslide hazard mapping, and it is expected that some suggestions in this study are helpful to further real applications including integration, and interpretation stages in order to obtain a decision-supporting layer.

## 1. INTRODUCTION

Landslides cause extensive damage to property and occasionally result in loss of life throughout most of country. So it is necessary to delineate the area that will be likely to be affected by the future landslides. For landslide susceptibility analysis, a unified and general framework has been proposed and termed Favourability Function by Chung and Fabbri (1993). FF models can be based on probability, evidential reasoning or fuzzy set theory, depending on the quantitative relationships between input causal factors and the past landslides. These approaches with their own mathematical backgrounds have provided powerful schemes for decision-supporting information, through several case studies (Van Westen, 1993; Chung and Fabbri, 1998, 1999; Carrara *et al*., 1998; Jibson *et al*., 1998; Lee and Min, 2001). Conventional probabilistic approaches implicitly assume that most of the information on which decision-making is based is probabilistic in nature, and that precise probability judgements can be formulated for each hypothesis of the problem concerned. On the other hand, in terms of soft computing, uncertainty may have different nature and should be modelled in different frameworks, and a hard decision should be drawn only towards the end of the processing (Binaghi *et al*., 1998). Especially, fuzzy set theory can provide us with a natural method of quantitatively processing multiple data sets and many scientists have applied the fuzzy set theory to their studies and proved that this theory is very useful to reflect natural phenomena or irregular behaviros (Zadeh, 1965; An *et al*., 1991; Chung and Fabbri, 1993; Zimmermann, 1996).

In this paper, we apply and investigate the fuzzy logic information representation and integration for landslide hazard mapping. First, we construct the input causal factors related to landslide occurrences, and then assignment of fuzzy membership functions is followed. To integrate fuzzy membership functions, we construct "fuzzy inference network" by using various fuzzy operators. As an essential part for landslide hazard mapping, in order to validate the significance of the prediction results, we exemplify whether and to what extent a prediction can be extended, in space, to neighbouring areas with similar geology. A case study from Boeun, Korea is carried out to illustrate above schemes.

## 2. STUDY AREA AND DATA SETS

The Boeun area in Korea, which had much landslide damage following heavy rain in 1998, was selected as the test area (Figure. 1). A two-day intensive rainfall between August 11 and 12, 1998 had induced many landslides in the study area. Landslides usually induced due to rainfall, local downpour, earthquakes and volcanic activities. Landslides triggered by heavy rainfall are the most common throughout Korea. Landslides are usually categorized into falls, topples, slides, spreads, and flows. Shallow landslides occur in material defined as engineering soils: unconsolidated, inorganic mineral, residual, or transported material (colluviums or alluviums) including rock fragments (Varnes, 1978). In the study area, the landslides were mainly debris flows that occurred during 3–4 hours of high intensity rainfall, or shortly after (Kim *et al*., 2000).
The input data for a test consist of several layers of map information (Table 1). The slope and aspect were calculated from the 1: 5,000 scale DEM. As for the soil data sets, the texture, topography, drainage, material, and thickness of soil were acquired from 1:25,000 scale soil maps. As for the forest data sets, the type, diameter, age, and density of timber were acquired from 1:25,000 scale forest maps. The lithology map was obtained from 1:50,000 scale geological map. After pre-processing, all data sets were built on a cell-based database, and

* Corresponding author.

the whole study area consists of 1,879 X 1,444 pixels (= 2,713,276 pixels), covering approximately 68km$^2$. Each pixel corresponds to a 5m by 5m area on the ground.

The aerial photographs taken in 1996 and 1999 were used to detect landslide locations, and the locations were verified by fieldwork. In total, 375 debris flows type landslides were mapped. The target pattern, with the entire landslide bodies, consists of two separate and distinct sub-areas, the scarp area and the deposit area. The geomorphologic characteristics of these two sub-areas are distinctly different. In this study, the topographically highest 20% of the scars of the landslides are considered as trigger areas.


(a) (b)

Figure 1. (a) The location map of study area, (b) photos showing one of landslide scars occurred in this area

| Data types | |
|---|---|
| Topographic data sets | Slope |
| | Aspect |
| Forest data sets | Forest type |
| | Forest diameter |
| | Forest age |
| | Forest density |
| Soil data sets | Soil texture |
| | Soil topography |
| | Soil drainage |
| | Soil material |
| | Soil thickness |
| Lithology | |
| Landslides location map | |

Table 1. Data sets used in this study

## 3. FUZZY INFORMATION REPRESENTATION

### 3.1 Fuzzy membership function

The fuzzy set theory was introduced by Zadeh (1965), which facilitates analysis of non-discrete natural processes or phenomena as mathematical formulae (Zimmermann, 1996).

If X={$x$} denotes a universe of the attribute values, the fuzzy set A in the X is the set of ordered pairs

$$A = \{x,\ \mu_A(x)\},\ x \in X \quad (1)$$

$\mu_A(x)$ is known as grade of membership of $x$ in the A. Usually, $\mu_A(x)$ is an integer or a floating number in the range [0,1] with 1 representing full membership and 0 non-membership. The grade of membership reflects a kind of ordering that is not based on probability but on admitted possibility. The value of $\mu_A(x)$ for the attribute value $x$ in A can be interpreted as the degree of compatibility of the predicate associated with set A and attribute value $x$.

Fuzzy membership functions closely associated with semantic analysis can be determined either normatively or empirically. The derivation of membership functions is crucial in fuzzy information processing and the lack of simple and generally acceptable methods to build membership functions may cause it less favourably with other information processing methods. Despite the lack of scientific foundation for membership functions, many fuzzy systems have demonstrated satisfactory performance when compared with two-valued logic system composed of crisp set theory (Zimmermann, 1996). The normative approach is commonly used for deriving membership functions for linguistic values because impreciseness inherent to these values is subjective. However, this approach was basically designed for engineering applications. So it is judged not suitable for geoscientific applications such as mineral potential mapping and landslide hazard mapping. For integration of multiple geological data sets, An *et al.*(1991) assigned the fuzzy membership functions using empirical procedure based on expert's opinion and the mineral deposit model. Most of studies for mineral potential mapping, assignment of fuzzy membership functions are based on expert's opinion. Meanwhile, for landslide hazard mapping, Chung and Fabbri (2001) assigned the fuzzy membership functions using the relationships between input causal factors and known past landslides.

### 3.2 Assignment of fuzzy membership function

In this study, assignment of fuzzy membership functions to each data layer followed Chung and Fabbri (2001)'s approach. Our target proposition is "a pixel *p* in the study area will be affected by future debris flow type landslides".

First, we investigated the relationships between input causal factors and past landslides. For this, the likelihood ratio function of each map, which can highlight the difference between areas affected by past landslides and areas not affected by past landslides, was calculated and compared with each other. In slope map, the steeper the slope, the greater the landslide possibility. Most landslides occurred between 15$^o$ and 35$^o$. The slope angle is an essential component of landslide susceptibility. In general, it is expected that low slope angles have a low possibility of landslides due to lower shear stresses associated with low gradients. Steep natural slopes, however, may not be susceptible to shallow landslides (Lee and Min, 2001). In aspect map, the landslide occurrence possibility value was similar at all directions. In forest maps, the possibility of landslide occurrence is higher in larch and artificial Chestnut trees, very small diameters, younger timber, and loose density forest. These results are related to location of forest and amount of roots. In soil maps, the possibility of landslide occurrence is higher in well-drained soil, red-yellow podzolic soils and lithosols, acidic rocks residuum, mountainous areas, thick soils. These results are related to increase of unit weight and shear stress of soil due to pore-water increase. In lithology map, most landslides had occurred in biotite granite areas.

The definition of membership functions for input causal factors was performed by using above relationships between landslides and input causal factors based on the likelihood ratio functions and slightly modified by considering expert's opinion.

## 4. DATA INTEGRATION

### 4.1 Fuzzy operators

Each data layer of target information denoted from fuzzy theory can now be integrated by using fuzzy operators. When two membership functions $\mu_A(x)$ and $\mu_B(x)$ are combined, Some of the useful fuzzy set operators are as follows (An *et al*., 1991; Chung and Fabbri, 2001):

1. Fuzzy OR

$$\mu_{OR}(x) = \text{MAX}\,[\,\mu_A(x)\,,\ \mu_B(x)\,] \quad (2)$$

2. Fuzzy AND

$$\mu_{AND}(x) = \text{MIN}\,[\,\mu_A(x)\,,\ \mu_B(x)\,] \quad (3)$$

3. Fuzzy Algebraic Sum

$$\mu_{SUM}(x) = 1 - \prod_{i=1}^{2} \mu_i(x) \quad (4)$$

4. Fuzzy Algebraic Product

$$\mu_{PRODUCT}(x) = \prod_{i=1}^{2} \mu_i(x) \quad (5)$$

5. Fuzzy $\gamma$ operator

$$\mu_\gamma(x) = [\mu_{SUM}(x)]^{\gamma} \times [\mu_{PRODUCT}(x)]^{1-\gamma} \quad (6)$$

When the fuzzy OR and AND operators are used, only one of the contributing fuzzy sets has an effect on the resultant value. The fuzzy algebraic sum and algebraic product operators make the resultant set larger than, or equal to the maximum value and smaller than, or equal to the minimum value among all fuzzy sets, respectively.

Meanwhile, the resultant set that is combined by the fuzzy $\gamma$ operator has the value between that of the fuzzy algebraic product operator and that of the fuzzy algebraic sum operator. The determination of optimum value is closely associated with degree of compensation between the two extreme confidence levels. In cases of $\gamma$ = 1 (full compensation) or $\gamma$ = 0 (no compensation), these operators with different values are equivalent to algebraic sum operator or production sum operator, respectively. Therefore, the choice of can produce the resultant value that can ensure a flexible compromise (Figure 2).

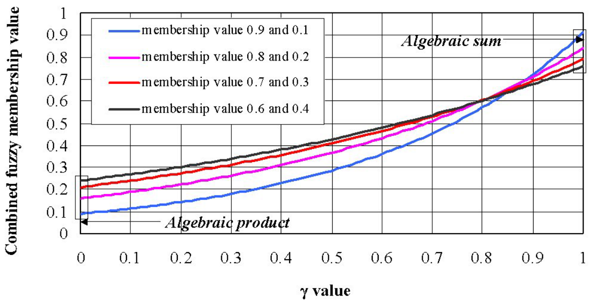


Figure 2. A graphical description of the resultant fuzzy membership value obtained by combining two fuzzy membership functions using fuzzy $\gamma$ operator

### 4.2 Design of fuzzy inference network

Though several authors have recently used fuzzy logic approaches for integration of multiple data sets, selecting an optimum fuzzy operator has always been a difficult task. It has become apparent that fuzzy operators depend very much on the types of spatial data to be integrated (Choi *et al*., 2000). In geoscientific applications of spatial data integration, spatial data have, in most cases, varying degree of information content with respect to the target proposition. In these cases, it is necessary to combine spatial data using several different fuzzy operators separately or a combination of selected operators depending on the characteristics of each data layer (Moon, 1998).

In this study, instead of using one operator, we constructed "fuzzy inference network" by using a variety of different fuzzy operators. We combined the fuzzy membership functions using the intermediate fuzzy information representation and various fuzzy operators (Figure 3). The intermediate fuzzy information representation is divided into three parts; topographic data sets, forest data sets, and soil data sets. The fuzzy OR operator was used in order to combine the topographic data sets including slope and aspect. The relationship between slop and aspect is not fully known and one may be considered as a primary information and the other is a secondary information. So we combined these maps using OR operators. When fuzzy OR operator is used, only one of the contributing fuzzy sets has an effect on the resultant set. While, for the intermediate fuzzy information representation of the forest data sets and the soil data sets, $\gamma$ operator was used to integrate them. Forest data sets and soil data sets have the typical characteristics wit respect to landslide occurrences. So certain classes of each map have a positive potential for landslide occurrences. However, the relationships among forest data sets and soil data sets are very complicated, so we would not expect that certain type of data have higher possibility than others. To consider this, fuzzy $\gamma$ operator, which can make it possible that all the contributing fuzzy sets have an effect on the resultant set, was used. Finally, to integrate three intermediate fuzzy information and lithology data, $\gamma$ operator is experimented.

We divided several fuzzy inference networks into 4 classes, depending on the choice of $\gamma$ value. Class 1 is one that all high $\gamma$ values were used for fuzzy intermediate representation and integration. In class 2, high $\gamma$ values were used for fuzzy intermediate representation of forest and soil data, and low or middle $\gamma$ values were tested for final integration. Class 3 is the opposite to class 2, that is, low or middle $\gamma$ values, and high $\gamma$ values were tested for fuzzy intermediate representation and integration, respectively. In class 4, all low $\gamma$ values were used.

Through above procedures, we prepared some prediction maps. To visualize the prediction maps, we used rank order statistics. We first computed the score for each pixel and then sort all scores by 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 maximum score has the maximum rank. Then the ranks are normalized so that the maximum value is 1 or 100%, and the normalized values are termed the favourability indices or simply indices. The pixel with the index 100% had the largest score of the prediction function. If the pixels have index, 99.5%, it means that the ranks of their function scores are within the top 0.5% (99.5% - 100%)

in the study area. These indices over the study area constitute a landslide susceptibility map.

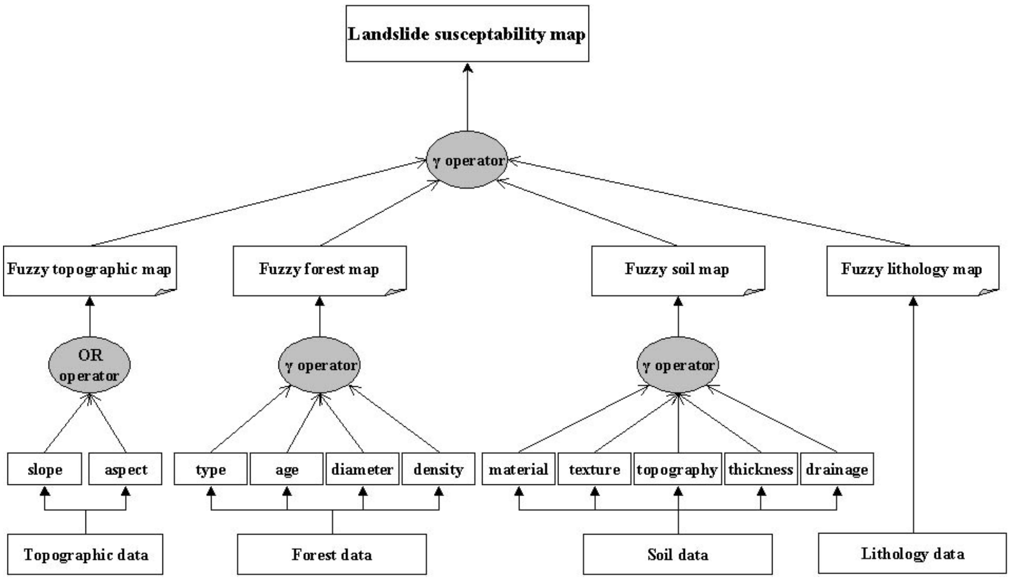


Figure 3. Fuzzy inference network designed in this study

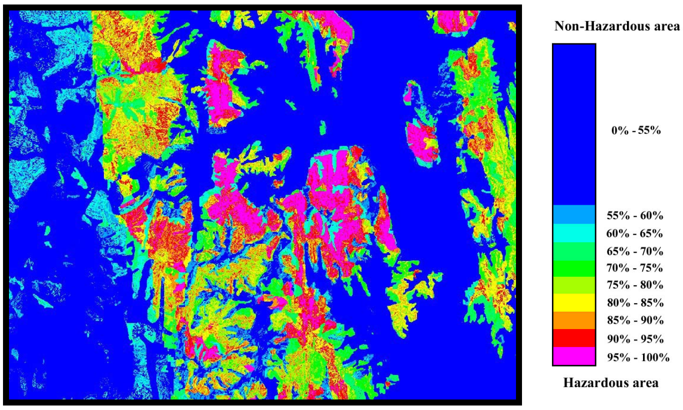


Figure 4. Prediction map using $\gamma = 1$ for intermediate fuzzy representation and $\gamma = 0.85$ for final integration

Figure 4 shows one example of fuzzy inference network using $\gamma = 1$ for forest data set and soil data set, $\gamma = 0.9$ for overall combination.
After we get the prediction maps, the most important question is "how successful this prediction map would be with respect to the future landslides". To answer this question, we need the information interpretable with respect to the future event. It leads to the next essential step of validation.

## 5. VALIDATION OF PREDICTION RESULT

The critical strategy in prediction models is the task of validating the prediction results, so that the prediction results can provide meaningful interpretation with respect to the future landslides (Fabbri and Chung, 2001; Chung and Fabbri, 2002). To carry out the validation, we must restrict the use of all the data of the past landslides in the study area. By partitioning the data, one subset is used for obtaining a prediction map; the other subset is compared with the prediction results for validation. To establish whether and to what extent a prediction can be extended, in space, to neighbouring areas with similar geology, we divided the entire study area into two separated sub-areas. The study area has been subdivided into a northern sub-area and a southern sub-area. This was because greater similarity exists between north-south than east-west sub-areas. We selected one of two sub-areas to construct a prediction model and the other to validate the prediction. Through this validation procedure, we can assess the prediction powers of various fuzzy inference networks, and compare with them quantitatively.
The space-partition technique used in this study consisted of the following steps (Figure 5):

- The 237 scarps distributed in the north sub-area were used to compute the south sub-area fuzzy inference networks.
- Similarly, the 138 scarps in the south sub-area were used to compute the north sub-area fuzzy inference networks.
- Then we assembled them into a mosaic of the two representations.

In order to validate a mosaic prediction map, we computed the prediction rate curve, which can explain the proportion of pixels correctly classified for the whole scarps in a mosaic map. This prediction rate curve relates to the number of the future landslides and to the probability of the occurrences of the future landslides.

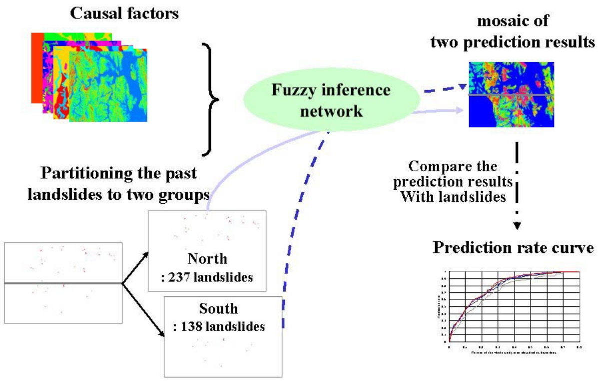


Figure 5. A graphical representation of cross-validation approach in this study

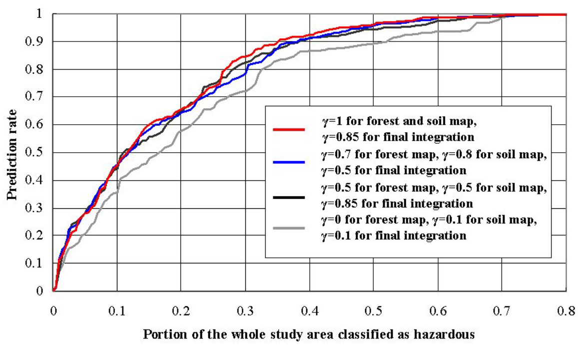


Figure 6. Prediction rate curves for 4 fuzzy inference networks

The prediction rate curves are shown in Figure 6. When we compare with the prediction powers among the four classes mentioned in section 4, the prediction powers of class 1 (red line in Figure 6), class 2 (blue line in Figure 6), and class 3 (black line in Figure 6) show the similar ones. In contrary to those, class 4 shows poorer prediction powers. For the future landslides, if we take the most hazardous 10% area of prediction images generated by various fuzzy inference networks, then we may estimate that about 45% of all future landslides will be located in the delineated area of class 1, class2, and class 3. The prediction rate curves are shown in Figure 6. When we compare with the prediction powers among the four classes mentioned in section 4, the prediction powers of class 1 (red line in Figure 6), class 2 (blue line in Figure 6), and class 3 (black line in Figure 6) show the similar ones. In contrary to those, class 4 shows poorer prediction powers. For the future landslides, if we take the most hazardous 10% area of prediction images generated by various fuzzy inference networks, then we may estimate that about 45% of all future landslides will be located in the delineated area of class 1, class2, and class 3. However, in case of class 4, about 35% of all future landslides will be located. In our study area, if we choose the relatively higher $\gamma$ value for integration, we may say that the results are effective and the fine difference of $\gamma$ value does not affect the final prediction results. When the relationship between data sets is not fully known and it is difficult to inference it, we would conclude that no compensation ( $\gamma$ =0) operator or small compensation may be inappropriate relatively.

## 6. DISCUSSION AND CONCLUSION

In this study, we applied fuzzy logic integration appraoch for landslide hazard mapping using multiple spatial data sets, and outlined the areas that will be affected by future landslides.

To combine various spatial data, fuzzy inference networks using combining some fuzzy memberships in series and others in parallel. Also, fuzzy $\gamma$ operators with various $\gamma$ values were tested. During the data representation and integration, fuzzy OR operator and $\gamma$ operator with high $\gamma$ value could effectively integrate most data sets. When we cannot be sure of relationships between multiple spatial data sets, fuzzy OR operator and $\gamma$ operator with high $\gamma$ value can be more effective than fuzzy $\gamma$ operator with no compensation or small compensation. However, we remind that the results in this study are not general ones, so extensive experiments should be made in several study areas to strengthen the situation here identified.

To assess quantitatively the prediction powers of various fuzzy inference networks, cross-validation approach was also performed. With the help of cross-validation approach, we can evaluate the prediction results quantitatively, and compare with models. Without this kind of the cross-validation technique, prediction maps cannot be evaluated.

For the future works, several aspects still need to be considered. For any prediction models to generate reasonably “good or significant” results, the prediction result should be robust and stable (Chung *et al*., 2002). For this, we are currently evaluating the stability analysis using matching rate function. In addition, we will try to involve the fuzziness of boundaries in categorical maps such as forest, soil, and lithology maps, in data representation stage.

## REFERENCES


An, P., Moon, W.M. and Rencz, A., 1991. Application of fuzzy set theory to integrated mineral exploration. *Canadian Journal of Exploration Geophysics*, 27(1), pp.1-11.

Binaghi, E., Luzi, L., Madella, P., Pergalani, F. and A. Rampini, 1998. Slope instability zonation: a comparison between certainty factor and fuzzy Dempster-Shafer approaches. *Natural Hazard*, 17, pp.77-97.

Carrara, A., Guzzetti, F., Cardinali, M. and Reichenbach, P., 1999. Use of GIS technology in the prediction and monitoring of landslide hazard. *Natural Hazards*, 20, pp.117-135.

Choi, S.W., Moon, W.M. and Choi, S.G., 2000. Fuzzy logic fusion of W-Mo exploration data from Seobyeog-ri, Korea. *Geosciences Journal*, 4(2), pp.43-52.

Chung, C. F. and Fabbri, A. G., 1993. The representation of geoscience information for data integration. *Nonrenewable Resources*, 2(2), pp.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 Annual*

*Meeting (IAMG'98)*, Ischia, Italy, October 3-7, 1998, pp.204-211.

Chung, C. F. and Fabbri, A. G., 1999. Probabilistic prediction models for landslide hazard mapping. *Photogrammetric Engineering & Remote Sensing*, 65(12), pp.1389-1399.

Chung, C. F. and Fabbri, A. G., 2001. Prediction models for landslide hazard zonation using a fuzzy set approach. . In, M. Marchetti, V. Rivas, ed., *Geomorphology & Environmental Impact Assessment*, Balkema Publishers, pp.31-47.

Chung, C.F. and Fabbri, A.G., 2002. Validation of spatial prediction models for landslide hazard mapping. *Natural Hazard* (in press).

Chung, C.F., Kojima, H., and Fabbri, A.G., 2002. Stability analysis of prediction models for landslide hazard mapping. In, R.J. Allison, ed., *Applied Geomorphology: theory and practice*, John Wiley & Sons, Ltd (in press).

Fabbri, A.G. and Chung, C.F., 2001. Spatial support in landslide hazard predictions based on map overlays, *Proceedings of International Association for Mathematical Geology Annual Meeting (IAMG'2001)*, Cancun, Mexico, September 10-12, 2001, CDROM.

Jibson, R.W., Harp, E.L., and Michael, J.A., 1998. A method to produce probabilistic seismic landslide hazard maps. In, A. Bucciantti, ed., *Proceedings of International Association for Mathematical Geology Annual Meeting (IAMG'98)*, Ischia, Italy, October 3-7, 1998, pp.211-217.

Kim, K.S., Kim, W.Y., Chae, B.G. and Cho, Y.C., 2000. Engineering geologic characteristics of landslide induced rainfall –Boeun, Chungcheong Buk-do-. The Korean Journal of Engineering Geology, 10(2), pp.163-174.

Lee, S. and Min, K.D., 2001. Statistical analysis of landslide susceptibility at Yongin, Korea. *Environmental Geology*, 40(9), pp.1095-1113.

Moon, W.M., 1998. Integration and fusion of geological exploration data: a theoretical review of fuzzy logic approach. *Geosciences Journal*, 2(4), pp.175-183.

Van Westen, C.J., 1993, *Application of geographic information systems to landslide hazard zonation,* Ph.D. Thesis, Technical university of Delft, International Institute for Aerospace Survey and Earth Sciences, The Netherlands, ITC publication 15(1), 245p.

Varnes, D.J., 1978. Slope movement types and processes, landslides analysis and control. *Special Report 176. Transportation Research Board*, Washington, DC pp.11-80.

Zadeh, L.A., 1965. Fuzzy sets. *IEEE Information and Control*, 8(3), pp.338-353.

Zimmermann, H.J., 1996. *Fuzzy set theory and its applications*. Kluwer Academic Publisher, 435p.