

산사태 취약성 분석을 위한 GIS 기반 확률론적 추정 모델과 모수적 모델의 적용

박노욱^{1*} · 지광훈¹ · Chang-Jo F. Chung² · 권병두³

¹한국지질자원연구원 지질자원정보센터, ²Geological Survey of Canada, ³서울대학교 지구과학교육과

Application of GIS-based Probabilistic Empirical and Parametric Models for Landslide Susceptibility Analysis

No-Wook Park^{1*}, Kwang-Hoon Chi¹, Chang-Jo F. Chung² and Byung-Doo Kwon³

¹Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources, Daejeon 305-350, Korea

²Geological Survey of Canada

³Department of Earth Sciences, Seoul National University, Seoul 151-742, Korea

Traditional GIS-based probabilistic spatial data integration models for landslide susceptibility analysis have failed to provide the theoretical backgrounds and effective methods for integration of different types of spatial data such as categorical and continuous data. This paper applies two spatial data integration models including non-parametric empirical estimation and parametric predictive discriminant analysis models that can directly use the original continuous data within a likelihood ratio framework. Similarity rates and a prediction rate curve are computed to quantitatively compare those two models. To illustrate the proposed models, two case studies from the Jangheung and Boeun areas were carried out and analyzed. As a result of the Jangheung case study, two models showed similar prediction capabilities. On the other hand, in the Boeun area, the parametric predictive discriminant analysis model showed the better prediction capability than that from the non-parametric empirical estimation model. In conclusion, the proposed models could effectively integrate the continuous data for landslide susceptibility analysis and more case studies should be carried out to support the results from the case studies, since each model has a distinctive feature in continuous data representation.

Key words : spatial integration, likelihood ratio, discriminant analysis, landslide susceptibility

산사태 취약성 분석을 위해 적용된 기존 GIS 기반 확률론적 공간 통합 모델은 범주형과 연속형 자료와 같이 서로 다른 형태의 자료의 처리를 위한 이론적 배경과 효율적인 방법론을 제시하지 못하였다. 이 논문에서는 우도비의 틀 안에서 연속형 자료를 직접적으로 사용할 수 있도록 비모수적 경험적 추정 모델과 모수적 예측적 판별 분석 모델을 적용하였다. 그리고 유사율과 예측비율곡선을 계산함으로써 두 모델을 정량적으로 비교하고자 하였다. 제안 모델을 비교하기 위해 1998년 여름 산사태로 많은 피해를 입은 장흥 지역과 보은 지역을 대상으로 사례연구를 수행하였다. 장흥 지역에서는 두 모델이 유사한 예측 능력을 나타내었으나, 보은 지역에서는 모수적 예측적 판별 분석 모델이 보다 높은 예측 능력을 나타내었다. 결론적으로 제안한 두 모델은 산사태 취약성 분석을 위한 연속형 자료 표현에 효율적으로 적용될 수 있으며, 두 모델이 개별적인 연속형 자료 표현의 특성을 가지고 있기 때문에 다른 사례 연구를 통한 검증 작업이 병행되어야 할 것으로 생각된다.

주요어 : 공간 통합, 우도비, 판별분석, 산사태 취약성

1. 서 론

해마다 여름철에 집중호우나 태풍으로 인한 산사태

는 인명과 재산 피해뿐만 아니라 국가 기반시설의 기능에도 피해를 입히고 있으며, 이러한 피해는 계속 증가되고 있는 상황이다(이사로, 2000; 김원영 외,

*Corresponding author: nwpark@kigam.re.kr

2003). 이러한 산사태 재해는 비록 강우나 지진 등의 요인에 의해 직접적으로 발생하지만, 다른 지구과학 현상과 마찬가지로 하나의 요인에 의해 발생한다기보다는 여러 요인이 복합적으로 작용하여 발생하게 된다. 따라서 산사태 피해를 분석하거나 예방하기 위해서는 여러 요인을 함께 분석할 수 있는 방법론의 개발이 필요한 상황이다.

산사태 발생과 관련된 공간자료를 효율적으로 다루기 위해 GIS의 적용가능성에 대한 연구가 1990년대부터 진행되어 왔다(김윤종 등, 1992, 1993; 한대석 등, 1998). 그러나 GIS가 레이어 개념을 도입함으로써 복합 자료의 효율적인 처리를 위한 환경을 제공하기는 하지만, 상업용 GIS 툴에서 제공하는 공간 분석 모듈은 대부분 단순 중첩 분석이나 주관적인 가중치 부여 등으로 지질학적 응용에는 한계가 있다.

다양한 지질자료를 효율적으로 통합 분석하기 위한 방법론 개발은 GIS의 발전과 더불어 혹은 개별적으로 확률/통계, 인공지능 등에 이론적 배경을 두고 1980년대 말부터 진행되어 왔다. 확률론적 모델로, Chung and Fabbri(1998)은 베이지안 이론하에 조건부 확률 모델(conditional probability model), 우도비 모델(likelihood ratio model), Certainty Factor 모델, Weight of Evidence 모델을 산사태 취약성 분석에 적용하였다. 최근에는 경기도 장흥 지역의 산사태 취약성 분석을 목적으로 Lee *et al.*(2002)이 우도비의 변형인 Weight of Evidence 모델을, Park *et al.*(2003)이 우도비 모델을 각각 적용하였다. 특히 대상 사건이 특정 속성에서 발생하는 상황만을 고려하는 조건부 확률 모델에 비해, 우도비 모델은 다른 속성에서 대상 사건이 발생하지 않은 상황에 대한 고려를 하기 때문에 다른 속성과의 상대적인 의미를 내포하는 장점이 있다. 그러나 범주형 자료와 연속형 자료를 함께 분석에 사용하는 기존 산사태 취약성 분석 연구의 경우, 대부분 서로 다른 형태의 자료(범주형, 연속형 자료) 처리와 확률론적 모델을 적용할 때 필요한 가정(조건부 독립, 선형성 등)에 대해서 논의없이 곧바로 적용하여 왔다.

이 연구에서는 산사태 취약성 분석을 목적으로 우도비 기반의 모델을 기반으로 비모수적(non-parametric) 및 모수적(parametric) 추정 모델을 각각 적용하였다. 연속형 자료의 우도비 계산을 위해 적용한 2가지 모델은 경험적 추정 모델(empirical estimation)과 예측적 판별 분석모델(predictive discriminant analysis model)이다. 기존 연구와의 차별성은 이 2가지 모델들

을 연속형 자료와 범주형 자료를 구분하여 처리할 수 있는 틀 안에서 이론과 적용시 필요한 기본 가정을 제시하고 비교 연구를 수행하는데 있다. 1998년 8월 여름 집중호우로 인해 산사태 피해를 입은 경기도 장흥 지역과 충청북도 보은 지역의 사례연구를 통해 모델의 정량적 비교와 적용시 제안점에 대해서 논의하였다.

2. 우도비 기반 공간통합 모델

2.1. 기본 개념

기본적인 전처리 과정을 거쳐 GIS 데이터베이스로 구축된 n 개의 지질 및 지형 자료 $e_k(k=1, \dots, n)$ 가 있고, 통합 목표 를 “특정 화소 T_p 는 미래의 특정 유형의 산사태에 의해 영향을 받을 것이다”라고 할 때, 산사태 취약성 분석의 목적은 e_k 를 T_p 에 적합하도록 정량적으로 추정 및 표현하는데 있다.

자연 현상을 수치화하는 과정에서 단순화 혹은 모델화시킨 지질 및 지형 자료에 대해 정량적 모델을 적용하기 위해서는 기본적인 가정이 필요한데, 산사태 취약성 분석을 위해 다음 2가지 가정을 하였다. 첫 번째 가정은 연구지역에서 특정 원인(강우, 지진 등)에 의해 미래에 발생할 산사태의 유형은 과거 발생한 산사태의 유형과 같다는 것이다. 만약 연구지역에 여러 유형의 산사태가 발생하였을 경우에는 유형별로 산사태를 따로 구분하여 모델링을 수행하여야 한다. 두 번째 가정은 통합 모델링에 사용하는 지질자료들은 과거나 미래의 산사태 발생을 예측하는데 유용한 정보를 제공한다는 것이다. 여러 지질자료가 산사태 발생과 관련하여 제공하는 정보의 질이 보통 다른 경우가 많은데, 이러한 경우에는 민감도 분석(sensitivity analysis)을 통해 각 지질자료의 기여도를 추출할 수 있다.

위의 2가지 가정 하에 여러 지질/지형 자료를 통합할 수 있는 방법론을 적용하게 되는데, 이 논문에서 적용한 확률론적 우도비의 기본 개념은 다음과 같다. 산사태 취약성 분석의 최종 목표는 미래의 산사태에 대해 취약한 지역과 취약하지 않은 지역을 구분하는데 있다. 두 번째 가정에 따라 산사태에 발생 지역과 미발생 지역에서의 지질/지형 자료들의 빈도수 분포 혹은 확률 분포의 차이가 나타날 것이며, 이러한 차이는 산사태 발생지역과 미발생 지역에서의 확률의 비로 정량적으로 부각될 수 있다.

베이지안 확률 이론에 의거하여 $e_k(k=1, \dots, n)$ 가 주어졌을 때, 산사태가 발생할 확률 $\Pr(T_p|e_1, \dots, e_n)$ 과 산사태가 발생하지 않을 확률 $\Pr(\bar{T}_p|e_1, \dots, e_n)$ 의 비로

표현되는 승산비(odds ratio) $O(T_p|e_1, \dots, e_n)$ 는 아래와 같이 정의될 수 있다(Pearl, 1997).

$$O(T_p|e_1, \dots, e_n) = \frac{\Pr(T_p|e_1, \dots, e_n)}{\Pr(\bar{T}_p|e_1, \dots, e_n)} \\ = \frac{\Pr(T_p) \cdot \Pr(e_1, \dots, e_n|T_p)}{\Pr(\bar{T}_p) \cdot \Pr(e_1, \dots, e_n|\bar{T}_p)} \quad (1)$$

여기서 \bar{T}_p 는 “특정 화소 p는 미래의 특정 유형의 산사태에 의해 영향을 받지 않을 것이다”를 나타내며, $\Pr(T_p)$ 와 $\Pr(\bar{T}_p)$ 는 통합 목표 T_p 와 \bar{T}_p 의 사전 확률(a prior probability)을 각각 나타낸다.

사전 승산비(a prior odds) $O(T_p)$ 와 우도비 $L(e_1, \dots, e_n|T_p)$ 는 각각 다음과 같이 정의될 수 있으며, 승산비 $O(T_p|e_1, \dots, e_n)$ 는 이 두가지 항의 곱으로 표현된다(Pearl, 1997).

$$O(T_p) = \frac{\Pr(T_p)}{\Pr(\bar{T}_p)} \\ L(e_1, \dots, e_n|T_p) = \frac{\Pr(e_1, \dots, e_n|T_p)}{\Pr(e_1, \dots, e_n|\bar{T}_p)} \quad (2)$$

$$O(T_p|e_1, \dots, e_n) = O(T_p) \cdot L(e_1, \dots, e_n|T_p)$$

산사태 취약성 분석의 경우, 사전 확률 $\Pr(T_p)$ 는 화소의 위치와 상관없이 산사태 발생 화소수를 연구지역 전체 화소수로 나눈 값으로 주어지게 된다. 따라서 $O(T_p)$ 는 연구지역 전 화소에 대해서 동일한 값을 가지게 되며, 승산비 $O(T_p|e_1, \dots, e_n)$ 는 우도비 $L(e_1, \dots, e_n|T_p)$ 에 비례하게 된다. 산사태 취약성 분석에서 중요한 점은 고려하고 있는 연구지역 내에서의 화소간의 상대적인 취약성의 수준의 비교이기 때문에 이 논문에서는 앞으로 우도비만을 고려하여 논의하고자 한다.

2.2. 우도비 표현

n개의 자료로부터 우도비를 계산하기 위해서는 개별적인 자료의 우도비의 계산과 이들의 통합이 중요하다.

보통 산사태 취약성 분석의 경우 지질도, 토양도, 임상도 등의 범주형 자료와 사면 고도, 경사, 방위 등의 연속형 자료를 함께 사용하게 된다. 기존 우도비를 기반으로 산사태 취약성 분석을 수행한 경우(Lee and Min, 2001), 연속형 자료를 이진화하거나 몇 개의 클래스를 가지는 범주형 자료로 변환하여 사용하여 왔다. 그러나 이 경우 범주형화 과정에서 정보의 손실이 따르며, 이 과정이 주관적일 수 있다(박노옥 등, 2003). 따라서 이 연구에서는 범주형 자료와 연속형 자료를 구분하여 처리하고, 특히 연속형 자료의 원래 값을 그

대로 사용하고자 하였다.

n개의 자료가 i개의 범주형 자료 $x_h(h=1, \dots, j)$ 와 j개의 연속형 자료 $y_l(l=1, \dots, j)$ 로 이루어져 있다고 할 때($i+j=n$), 범주형 자료와 연속형 자료사이에 조건부 독립(conditional independence)을 가정하게 되면 개별적인 자료 처리가 아래와 같이 가능하게 된다(Spatial-Models Inc., 2004).

$$L(e_1, \dots, e_n|T_p) = L(x_1, \dots, x_i|T_p) \cdot L(y_1, \dots, y_j|T_p) \quad (3)$$

또한 조건부 독립의 가정을 범주형 자료 사이와 연속형 사이에 적용할 경우, 우도비는 개별적인 자료의 우도비의 곱으로 표현될 수 있다.

$$L(x_1, \dots, x_i|T_p) = L(x_1|T_p) \cdot L(x_2|T_p) \cdots L(x_i|T_p) \\ L(y_1, \dots, y_j|T_p) = L(y_1|T_p) \cdot L(y_2|T_p) \cdots L(y_j|T_p) \quad (4)$$

식 (4)와 (5)로부터 최종적인 우도비는 식 (5)으로 표현된다(Chung and Fabbri, 1998).

$$L(e_1, \dots, e_n|T_p) = L(x_1|T_p) \cdots L(x_i|T_p) \cdot L(y_1|T_p) \cdots L(y_j|T_p) \quad (5)$$

기존 이사로(2000)와 Lee and Min (2001)의 연구에서는 빈도수비를 이용하여 산사태 취약성 분석을 시도하였는데, 개별적인 자료의 빈도수비의 합으로 산사태 취약성 지수를 계산하였다. 그러나 빈도수비는 수식적으로 우도비와 일치하며, 엄밀한 의미에서는 조건부 독립의 가정하에 우도비를 곱해주어야 한다.

확률론적 관점에서 조건부 독립은 복잡한 여러 자료의 통합을 통계적으로 표현할 수 있는 형태로 변형하는데 필요한 가정에 해당된다. 조건부 독립의 가정은 수학적으로 다루기 편한 형태를 제공하는 하나, 실제 내부적으로 복잡한 관계를 가지는 자료들에 대해서 증명하기가 쉽지 않은 경우가 많다. Bonham-Carter (1994)는 조건부 독립을 정량적으로 판단하기 위한 테스트를 시도하여 조건부 종속인 자료를 제외할 것을 제안하였으나, 조건부 독립의 가정은 개념적인 의미가 크다. 만약 두 개의 서로 다른 자료가 공간적으로 유사한 양상을 보일 경우, 조건부 독립 테스트의 결과는 서로 조건부 종속(conditional dependence)으로 나타날 수 있다. 그러나 자료 통합의 관점에서 볼 때, 서로 다른 자료가 공간적으로 유사한 양상을 보인다고 하더라도 통합 목적에 대해 개별적으로 비슷한 정보를 제공한다고 간주할 수 있기 때문에 분석에 모두 사용해야 할 경우가 발생한다. 따라서 조건부 독립은 엄밀하게 말해서는 통합 목표에 대한 자료들의 공간적 분포

의 관계보다는 자료 사이의 개념적인 관계와 그 자료들이 얼마나 통합목표와 연관되어 있느냐와 관계된다.

식 (3), (4), (5)에서 범주형 자료의 우도비는 빈도수 분포로부터 계산될 수 있다. m 개의 속성을 가지는 범주형 자료 $x_h(a)$ ($a=1, \dots, m$)의 산사태가 발생할 확률 $\Pr(x_h(a)|T_p)$ 과 산사태가 발생하지 않을 확률 $\Pr(x_h(a)|\bar{T}_p)$ 및 우도비 $\text{Ph}(x_h(a)|T_p)$ 는 아래와 같이 정의된다.

$$\begin{aligned} \Pr(x_h(a)|T_p) &= \frac{N[x_h(a) \cap L]}{N[L]} \\ \Pr(x_h(a)|\bar{T}_p) &= \frac{N[x_h(a)] - N[x_h(a) \cap L]}{N[A] - N[L]} \\ L(x_h(a)|T_p) &= \frac{\Pr(x_h(a)|T_p)}{\Pr(x_h(a)|\bar{T}_p)} \end{aligned} \quad (6)$$

여기서 $N[\cdot]$ 은 \cdot 의 개수, $N[x_h(a) \cap L]$ 은 속성 a 에서 발생한 산사태의 개수, $N[L]$ 은 연구지역에서의 산사태 개수, $N[A]$ 은 연구지역 전체 화소수를 각각 나타낸다.

범주형 자료의 우도비 표현과 달리 연속형 자료는 원래 값을 그대로 표현할 수 있도록 빈도수 분포가 아닌 확률 분포를 이용하였다. 이를 위해 경험적 비모수적 추정 모델과 모수적 예측적 판별 분석을 적용하였는데, 각각의 방법에 대한 설명은 2.3과 2.4에 제시하였다.

2.3. 비모수적 경험적 추정 모델

비모수적 경험적 모델에서는 특정 확률 분포를 가정하지 않고, 자료 자체의 값으로부터 확률 분포를 추정하게 된다. 비모수적 확률 분포 추정 방법으로는 통계적 패턴 인식 분야에서 평활화 커널 추정(smoothed kernel estimation)과 k -nearest neighbor estimation 등이 대표적으로 사용되고 있는데(Duda *et al.*, 1997), 아직까지 지질자료의 표현에 사용된 연구는 거의 없는 상황이다. 이 연구에서는 비모수적 확률 분포 추정을 위해 평활화 커널 추정 방법(smoothed kernel estimation)을 이용하였다.

평활화 커널 방법은 샘플 자료에 커널 함수를 위치시킨 후 커널 함수의 중점으로 경험적 확률 밀도 분포를 계산하는 방법에 해당된다(Parzen, 1962). 산사태 발생 위치를 X_h , 산사태가 발생하지 않은 위치를 \bar{X}_h , 특정 연속형 자료를 y_i 이라고 할 때 가우시안 커널 함수를 이용하여 계산되는 산사태가 발생할 확률 $\Pr(y_i|T_p)$ 과 산사태가 발생하지 않을 확률 $\Pr(y_i|\bar{T}_p)$, 우도비 $L(y_i|T_p)$ 는 아래와 같이 정의된다.

$$\Pr(y_i|T_p) = \frac{1}{N(L)} \sum_{a=1}^{N(L)} \frac{1}{h\sqrt{2\pi}} \exp\left[-\frac{(y_i(X) - y_i(X_a))^2}{2h^2}\right] \quad (7)$$

$$\Pr(y_i|\bar{T}_p) = \frac{1}{N(A) - N(L)} \sum_{a=1}^{N(L)} \frac{1}{h\sqrt{2\pi}} \exp\left[-\frac{(y_i(X) - y_i(X_a))^2}{2h^2}\right]$$

$$L(y_i|T_p) = \frac{\Pr(y_i|T_p)}{\Pr(y_i|\bar{T}_p)}$$

여기서 $y_i(\cdot)$ 는 \cdot 위치에서의 속성값을, h 는 가우시안 커널의 폭을 결정하는 상수를 각각 나타낸다.

식 (7)을 통해 계산된 연속형 자료의 우도비와 식 (6)을 통해 계산된 범주형 자료의 우도비를 식 (5)을 통해 결합함으로써 최종적인 산사태 취약성 지수를 얻게 된다.

2.4. 모수적 예측적 판별 분석 모델

모수적 모델에서는 연속형 자료를 개별적으로 처리하는 것이 아니라 표본(산사태 취약성 분석의 경우 과거 산사태 발생 화소와 나머지 미발생 화소)으로부터 다변량 확률 분포를 가정하는 모수를 추정하여 분석을 수행하게 된다. 모수적 모델을 적용할 때 필요한 모수는 평균과 공분산이며, 표본 평균과 공분산을 모집단의 평균과 공분산의 추정치로 간주하게 된다.

이 논문에서 적용한 예측적 판별 분석 모델은 다변량 정규 분포를 가정하는 최대우도 추정 모델(maximum likelihood estimation model)과는 달리 다변량 student-t 분포를 가정하게 된다(McLachlan, 1992). 최대우도 추정 모델과의 차이는 표본 통계치가 얼마나 모집단 통계치에 근접한지에 대한 불확실성을 고려할 수 있다는 점이다. 표본수가 클 경우, student-t 분포는 정규 분포에 접근하여 두 모델의 결과가 유사하게 나타나지만, 표본수가 작은 경우에는 최대우도 추정 모델의 추정치를 조절할 수 있기 때문에 보다 나은 결과를 나타내는 것으로 알려져 있다(McLachlan, 1992).

예측적 판별분석에서 연속형 자료 y_i ($i=1, \dots, j$)에서 산사태가 발생할 확률 $\Pr(y_1, \dots, y_j|T_p)$ 과 산사태가 발생하지 않을 확률 $\Pr(y_1, \dots, y_j|\bar{T}_p)$, 우도비는 아래와 같이 정의된다(McLachlan, 1992; SpatialModels Inc., 2004).

$$\begin{aligned} \Pr(y_1, \dots, y_j|T_p) &= \left[\frac{N(L)}{(N(L)^2 - 1)} \right]^{j/2} \cdot \frac{\Gamma[N(L)/2]}{\Gamma[N(L) - j/2] |S_{T_p}|^{j/2} \pi^{j/2}} \\ &\cdot \left[1 + \frac{N(L)}{N(L)^2 - 1} (y - \hat{y}_{T_p})' \hat{S}_{T_p}^{-1} (y - \hat{y}_{T_p}) \right]^{-N(L)/2} \\ \Pr(y_1, \dots, y_j|\bar{T}_p) &= \left[\frac{N(A) - N(L)}{((N(A) - N(L))^2 - 1)} \right]^{j/2} \end{aligned}$$

$$\frac{\Gamma\left[\frac{N(A)-N(L)}{2}\right]}{\Gamma\left[\frac{N(A)-N(L)-j}{2}\right]\hat{S}_{\bar{y}_p}^{1/2}\pi^{1/2}} \quad (8)$$

$$\cdot \left[1 + \frac{N(A)-N(L)}{(N(A)-N(L))^2 - 1} (y - \hat{y}_{T_p})' \hat{S}_{\bar{y}_p}^{-1} (y - \hat{y}_{\bar{y}_p})\right]^{-(N(A)-N(L))/2}$$

$$L(y_1, \dots, y_j | T_p) = \frac{\Pr(y_1, \dots, y_j | T_p)}{\Pr(y_1, \dots, y_j | \bar{T}_p)}$$

여기서 \hat{y}_{T_p} , $\hat{y}_{\bar{T}_p}$, \hat{S}_{T_p} , $\hat{S}_{\bar{T}_p}$ 는 T_p 와 \bar{T}_p 의 평균과 공분산 추정치 벡터를 각각 나타내며, Γ 는 감마 함수에 해당된다.

식 (8)을 통해 계산된 연속형 자료의 우도비와 식 (6)을 통해 계산된 범주형 자료의 우도비를 식 (3)을 통해 결합함으로써 최종적인 산사태 취약성 지수를 얻게 된다.

3. 사례 연구

이 연구에서 제안한 우도비 기반 2가지 모델의 실제 적용과 비교를 위해 1998년 8월 여름에 집중호우로 많은 산사태 피해를 입은 경기도 장흥 지역과 충청북도 보은 지역을 대상으로 사례 연구를 수행하였다. 제안 모델의 적용을 위해 Fortran 프로그래밍과 Spatial-Models Inc.의 Spatial Prediction Modeling System을 이용하였다. 사례 연구를 통해 제안한 이론의 실제 적용과정에서의 적용 가능성과 모델의 비교에 대한 논의를 하고자 하였다.

3.1. 사용자료

경기도 장흥 지역 사례연구에는 범주형 자료로 임상종류도, 토양도를, 연속형 자료로 고도, 사면 방위, 사면 방위를 이용하였다. 정량적 비교를 위해 충청북도 보은 지역의 경우에도 동일한 자료를 이용하였다. 각 지역의 자료구축 과정과 다른 GIS 기반 공간 통합 모델의 적용 결과는 Park *et al.*(2003)과 박노옥 등 (2003)를 참조할 수 있다. 산사태 발생 개소수는 경기도 장흥 지역의 경우 332개, 충청북도 보은 지역의 경우 459개이며, 두 지역 모두 5미터 해상도의 GIS 공간 데이터베이스로 구축하였다(Fig. 1).

3.2. 통합 결과

제안된 두 모델의 차이는 연속형 자료의 표현에 있기 때문에, 연속형 자료의 표현과 통합에 주요점을 두

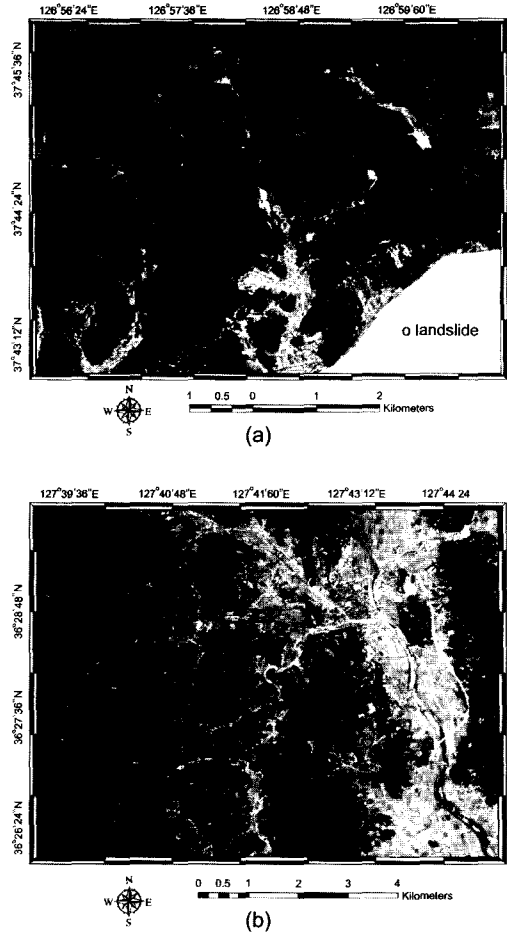


Fig. 1. KOMPSAT-1 EOC imagery with past landslide locations in the case study areas: (a) Jangheung, (b) Boeun.

고 사례연구를 수행하였다. 우선 범주형 자료의 우도비는 식 (6)을 이용하여 계산하여 분석에 사용하였다.

비모수적 경험적 추정 모델에서 평활화 커널의 폭을 결정하는 상수 h 는 실험적으로 전체 자료 범위의 4%로 정하였으며, 식 (7)을 이용하여 우도비를 계산하였다. 경험적 추정 모델의 경우, 개별적인 자료에 대한 확률 분포를 확인할 수 있기 때문에 두 지역에 대한 비교 분석을 수행하였다. 고도의 경우(Fig. 2 (a)), 두 지역 모두 특정 고도를 정점으로 우도비의 특성이 구별되는 것으로 나타났다. 우도비가 1보다 클 경우, 산사태 발생 확률이 높음을 나타내는데, 장흥 지역에서는 주로 고도 160 미터에서 280 미터에서, 보은 지역에서는 290 미터에서 480 미터에서 나타났다. 사면 경사의 경우(Fig. 2 (b)), 두 지역 모두 경사각이 증가할수록 우도비도 증가하는 양상을 보이고 있으나, 고도와

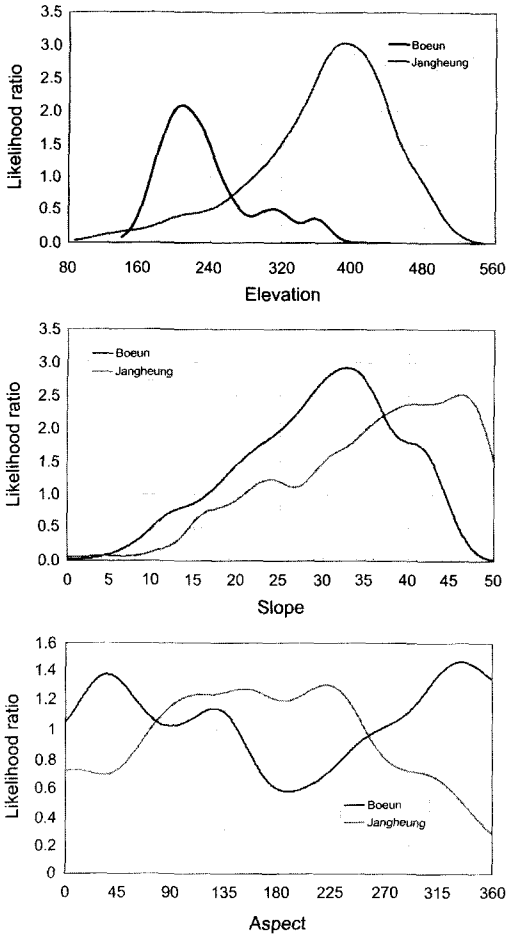


Fig. 2. Likelihood ratio values of (a) elevation, (b) slope, (c) aspect.

마찬가지로 특정 경사각을 정점으로 증가 추세가 감소하는 것으로 나타났다. 장흥 지역에서는 20도 이상에서 산사태가 주로 발생한 것으로 나타났으며, 경사각이 증가할수록 우도비가 증가하다가 48도 이상에서는 증가 추세가 감소하고 있다. 보은 지역에서도 대략 17도 이상에서 산사태가 주로 발생하였고, 증가 추세는 33도를 정점으로 해서 감소하는 것으로 나타났다. 이러한 결과는 일반적으로 고 경사각 지역에서는 표면 토양이 주로 침식되기 때문인 것으로 판단된다. 사면방위의 경우(Fig. 2 (b)), 장흥 지역에서는 동, 남동, 남, 남서쪽에서 산사태가 주로 발생하였고, 보은 지역에서는 반대로 북, 북동, 북서, 서쪽에서 주로 발생한 것으로 나타났다. 기존 연구에 의하면, 일반적으로 산사태 발생은 일반적으로 사면 방위로 인한 일조시간 차이에 의한 토층내의 지하수의 위치, 기반암의 풍화정도, 식생

의 생육차이로 인해 사면 방위가 남향인 것의 북경이 북향의 것보다 많다(토질안정재료위원회편, 1990). 그러나 급경사 지역이 선구조나 계곡 등에 의해 몇 개의 방향으로만 발달하여 경사각과 사면이 서로 연관되어 나타날 수 있기 때문에 남쪽 사면이 아닌 지역에서도 나타날 수 있다는 박용원 등(1993)의 연구와 연관이 있는 결과가 나타났다. 따라서 이사로(2000)에서 지적한 것처럼, 산사태와 사면방위에 대한 관계는 다른 연구지역과의 비교가 더 필요하다고 판단된다. 고도, 사면 경사 및 사면 방위의 전반적인 우도비의 크기를 비교해 보면, 두 지역 모두 고도와 사면경사의 최대 우도비 값이 크게 나타나고 있다. 사면 방위에서는 방위각에 따른 우도비의 차이가 크게 나타나고 있지 않다. 이러한 결과는 지형 자료 중에서 사면 방위에 비해 고도와 사면 경사가 산사태 발생과 미발생지역의 특성을 구별하는데 더 효과적임을 지시한다. 이렇게 계산된 개별적인 자료의 우도비를 식 (6)을 이용하여 연구지역의 산사태 취약성 분포도를 작성하였다(Fig. 3(a), Fig. 4(a)).

모수적 예측적 판별 분석에서는 우선 산사태 발생 지역과 미발생 지역의 평균과 공분산을 계산한 후, 식 (8)을 통해 우도비를 계산하였다. 이렇게 계산된 연속형 자료의 우도비와 범주형 자료의 우도비를 식 (3)을 통해 결합함으로써 연구 지역의 산사태 취약성 분포도를 작성하였다(Fig. 3(b), Fig. 4(b)).

Fig. 3과 Fig. 4에 제시된 두 지역의 산사태 취약성 분포도는 박노옥 등(2003)의 rank 기반 등면적 표현을 통해 작성되었다. 시각적으로 살펴보면, 두 모델의 전반적인 양상이 대체로 일치하고 있다. 즉 분홍색과 빨간색 계열로 나타나는 상대적 취약성이 높은 지역과 파란색 계열로 나타나는 상대적 취약성이 낮은 지역의 위치가 유사하게 나타나고 있다.

이러한 시각적 분석 이외에 각 모델의 유사성의 비교를 위해, 정량적 지수로 Chung *et al.*(2002)의 matching rate를 확장한 유사율(similarity rate: SR)을 정의하고 계산하였다. 유사율 계산을 위해, 우선 Fig. 3과 Fig. 4의 두 모델의 취약성 분포도에서 상위 취약성 등급 5%별로 특정 누적 등급을 산사태 취약지역으로, 나머지 비취약지역의 2개의 클래스로 구분을 하였다. 이러한 두 모델별로 구분된 2개의 클래스를 중첩할 경우, 공통 취약 클래스, 공통 비취약 클래스, 한 모델에서는 취약지역으로 구분되지만 다른 모델에서는 비취약지역으로 구분되는 클래스, 반대의 경우 등 모두 4개의 클래스를 얻게 된다. 이로부터 유사율 SR은

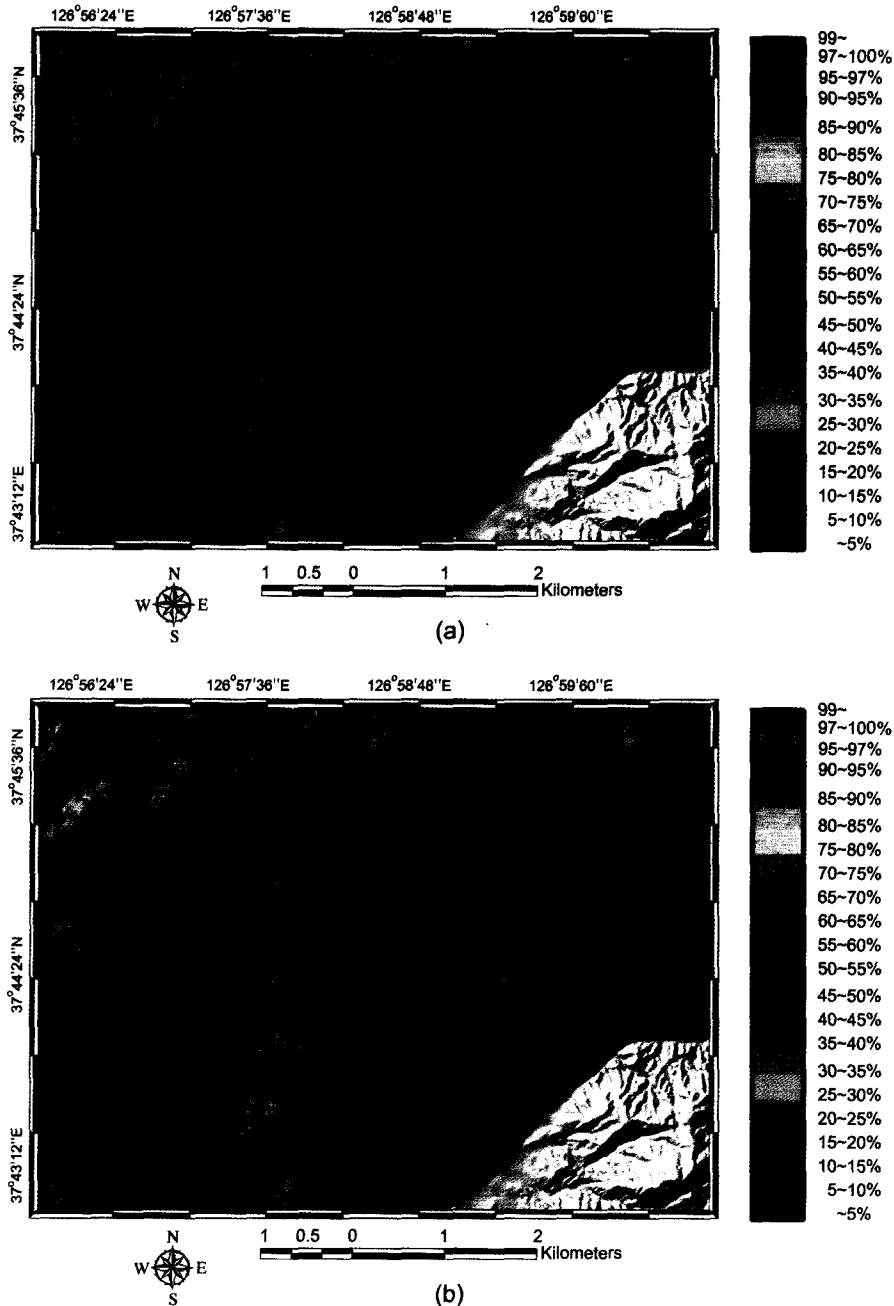


Fig. 3. Landslide susceptibility maps draped over the shaded relief map in the Jangheung area, (a) Empirical likelihood ratio estimation model, (b) Predictive discriminant analysis model.

두 모델의 공통으로 나타나는 산사태 취약지역의 화소 수를 계산한 후에, 이 값을 전체 화소수에서 공통 비 취약지역의 화소수를 뺀 값으로 나눠주어 계산하게 된다(식 (9)). 만약 두 모델의 취약성 등급의 위치가 유사할 경우, 공통 취약 지역과 비취약 지역의 화소수가

많이 나타나게 되어 유사율이 높아지게 된다.

$$SR = \frac{N[M_1(\text{susceptible areas}) \cap M_2(\text{susceptible areas})]}{N[A] - N[M_1(\text{non-susceptible areas}) \cap M_2(\text{non-susceptible areas})]} \quad (9)$$

여기서 M_1 과 M_2 는 비교할 공간통합 모델을, $N[\cdot]$

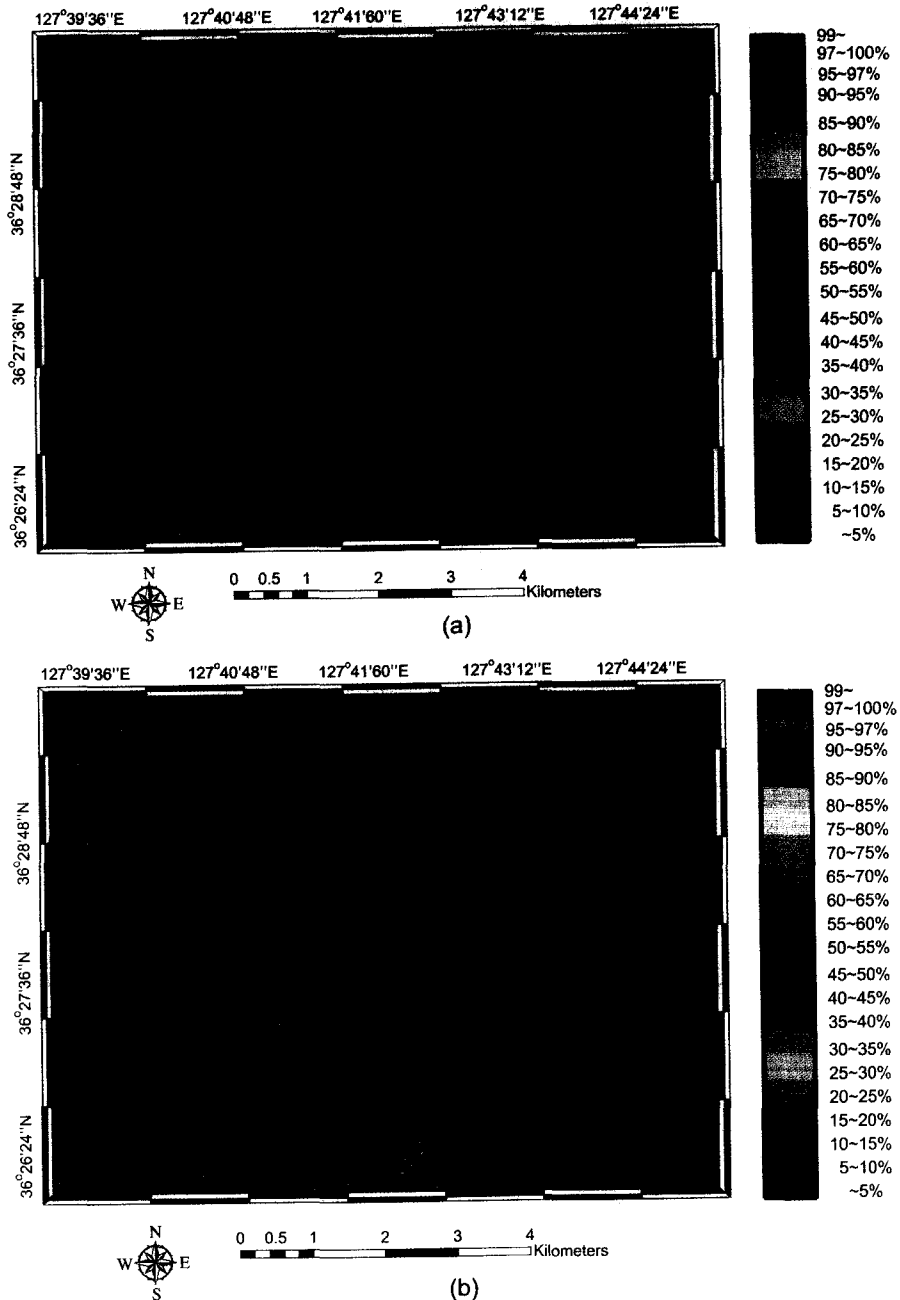


Fig. 4. Landslide susceptibility maps draped over the shaded relief map in the Boeun area. (a) Empirical likelihood ratio estimation model, (b) Predictive discriminant analysis model.

은 ·의 개수, $N[A]$ 는 연구지역 전체 화소수를 각각 나타낸다.

장흥과 보은 지역의 유사율을 살펴보면(Fig. 5), 전반적으로 장흥 지역이 보은 지역보다 높게 나타남을

확인할 수 있다. 상위 5%와 10% 취약성 등급을 살펴보면, 장흥 지역에서는 약 75%, 73%, 보은 지역에서는 약 52%, 65%로 각각 나타났다. 이는 장흥 지역에서는 비모수적 경험적 우도비 추정 모델과 모수적 예

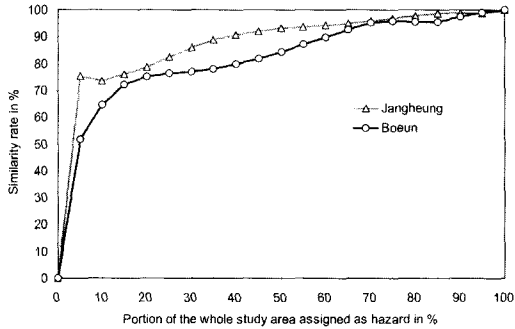


Fig. 5. Similarity rates for the Jangheung and Boeun areas.

측적 판별 분석 모델이 유사한 예측 패턴을 보이는 것으로 간주할 수 있다. 이러한 결과는 Fig. 3과 Fig. 4에 나타난 산사태 취약성 분포도의 시각적 비교만으로는 얻을 수 없는 정보에 해당되며, 이는 다음에 설명할 예측 능력의 비교와도 연관이 있다.

3.3. 검증 결과

각 모델의 예측 능력을 정량적으로 비교하기 위하여, 일종의 검증(validation)으로 과거 산사태를 분할하여 계산하는 예측비율곡선(Chung and Fabbri, 1999)을 작성하였다. 예측비율곡선은 과거 산사태의 분할에 기반하여 계산되는데, 이 연구에서는 임의적 공간 분할 방법을 통해 예측비율곡선을 작성하였다. 과거 산사태 발생위치를 임의의 분할을 통해 두 개의 그룹으로 나누어서 한 개의 그룹은 통합용, 나머지 그룹은 검증용에 사용하였다. 그리고 반대로 검증용 자료를 통합용 자료로, 그리고 통합용 자료를 검증용 자료로서 사용하였다. 최종적으로는 식 (10)을 이용하여 각 취약성 등급에 해당하는 산사태 발생 화소수를 계산하여 누적합수 형태인 예측비율곡선을 작성하였다(Chung and Fabbri, 1999).

$$PR(X_{\alpha}) = \frac{\sum_{i=1}^{\alpha} N[S_{100-X_i} \cap L]}{N[L]} \quad (10)$$

여기서는 취약성 등급에서의 누적예측비율을 나타낸다. 예를 들어 예측비율곡선에서 10%는 Fig. 3과 Fig. 4의 범례에서 상위 90% 취약성 등급(=100-10)에 해당된다. 그리고 은 취약성 등급에서 발생한 산사태의 개수, 은 연구지역에서의 전체 산사태 개수를 각각 나타낸다.

장흥 지역의 경우(Fig. 6 (a)), 두 모델의 예측 능력이 거의 유사하게 나타났다. 두 모델 모두 상위 10% 취약성 등급에서 약 45%의 예측 능력을 나타내고 있

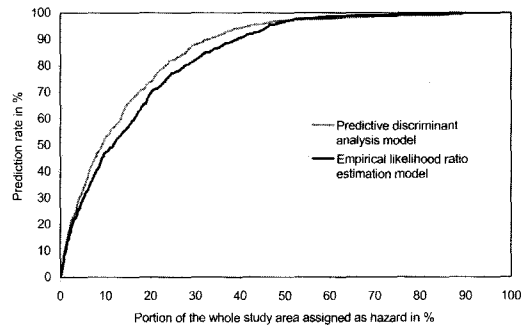
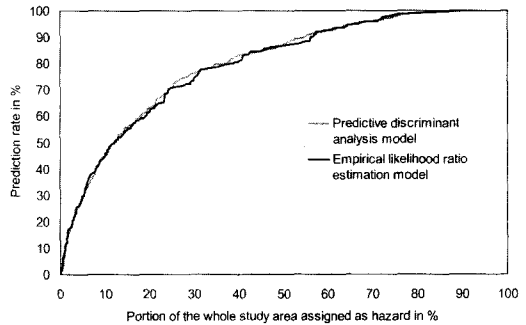


Fig. 6. Prediction rate curves for (a) Jangheung, (b) Boeun.

다. 두 모델간의 유사한 예측 능력은 앞 절에서 제시한 유사율과도 연관성이 있다. Fig. 5에서 장흥 지역의 경우 보은 지역에 비해 두 모델간의 유사율이 높게 나타나고 있다. 이러한 두 모델의 통합결과의 유사성에 의해 예측 능력 또한 유사하게 나타난 것으로 판단된다. 모델 구성의 이론적 관점에서 보면, 장흥 지역의 연속형 자료는 비모수적 추정과 모수적 추정을 통해 모두 적절하게 표현될 수 있다고 간주할 수 있다.

그러나 보은 지역의 경우(Fig. 6 (b))에서는 두 모델의 예측 능력의 차이가 나타나고 있다. 비모수적 우도비 추정 모델의 경우, 상위 10%에서 47%의 예측 능력을 나타내는 반면, 모수적 예측적 판별 분석 모델에서는 53%의 예측 능력을 보이고 있으며, 이러한 양상은 상위 50% 취약성 등급까지 계속되고 있다. Fig. 5에서 살펴 본 것처럼, 보은 지역에 비해 장흥 지역에 두 모델의 유사율이 현저하게 떨어지고 있다. 즉 두 모델의 통합 결과의 불일치 부분이 장흥 지역에 비해 많이 나타난다고 해석할 수 있다. 따라서 이러한 상이한 유사율이 최종적으로 교차 검증을 통해 예측 능력에도 반영되었다고 판단된다. 우도비의 이론적 틀 안에서 범주형 자료의 우도비는 두 모델에서 모두 동일하게 곱해지기 때문에, 연속형 자료의 영향에 의해 예측 능력이 좌우된다. 따라서 모수적 모델이 비모수적 모

델보다 예측 능력이 상대적으로 높은 것으로 나타난 결과는, 자료 표현의 관점에서 보든 지역의 연속형 자료가 모수적 다중 변수 표현으로 더 효과적으로 모델링 될 수 있다는 것으로 해석할 수 있다.

4. 토의 및 결론

이 논문에서는 산사태 취약성 분석을 위한 기존 GIS 기반 확률론적 공간 통합 모델에서 고려하지 않은 기본 가정을 포함한 이론적 틀을 제시하고, 사례연구를 수행하였다. 우도비를 기반으로 공간 통합 이론을 적용하였는데, 기존 우도비 기반 모델과 가장 큰 차별성은 연속형 자료의 직접적인 이용에 있다. 연속형 자료의 표현을 위해 비모수적 경험적 추정 모델과 모수적 예측적 판별 분석 모델을 적용하였으며, 유사율과 예측비율곡선의 계산을 통해 장흥 지역과 보은 지역의 사례연구 결과를 정량적으로 비교 분석을 수행하였다. 특히 서로 다른 모델을 적용하여 얻어진 공간 통합 모델의 공간적 분포를 정량적으로 비교하고자 유사율을 계산하였다.

사례연구 수행 결과, 장흥 지역에서는 두 모델의 유사율이 비슷하게 나왔으며 교차 검증을 통한 예측 능력 비교에서도 유사하게 나왔다. 반면 보은 지역에서는 두 모델의 유사율이 10% 취약성 등급에서 장흥 지역에 비해 8% 정도 낮게 나타났다. 또한 모수적 예측적 판별 분석 모델의 예측 능력이 비모수적 경험적 우도비 추정 모델보다 전반적으로 높게 나왔으며, 상위 10% 취약성 등급에서 6% 정도의 차이를 보였다. 범주형 자료의 우도비가 동일하게 작용하는 두 모델에서 예측 능력의 차이는 연속형 자료의 표현에 좌우되게 된다. 보은 지역처럼 모수적 모델이 비모수적 모델보다 예측 능력이 높게 나온 결과는 모수적 모델이 연구지역의 지형자료를 산사태 발생과 정량적으로 표현하는데 효과적임을 의미한다. 기존 GIS나 패턴 인식 분야에서는 지형 자료의 경우 다변량 모수적 모델링이 비모수적 모델링에 비해 비효율적이라고 보고하였으나, 보은 지역 사례 연구 결과는 반대로 나타났다. 비록 동일한 지형자료라 하더라도 연구 지역에 따라 특성이 달라지기 때문에, 보은 지역의 결과만으로 모수적 예측적 판별 분석 모델이 비모수적 경험적 추정 모델보다 적용 가능성이 높다고 결론지을 수는 없다. 자료 표현의 과정에서 비모수적 경험적 추정 모델은 개별 자료에 대한 우도비를 계산하기 때문에 개별 자료들의 산사태 발생과의 연관성 해석에 직접적으로 이용할 수 있는 장

점이 있다. 반면 모수적 예측적 판별 분석 모델은 다변량 분포의 파라미터로부터 우도비를 추정하기 때문에 개별 자료의 연관성은 평균과 공분산을 통해서만 해석할 수 있다. 따라서 이 연구를 통해 얻어진 결과를 일반화시키기 위해서는 보다 많은 사례 연구를 통한 검증 작업이 수행되어야 할 것으로 판단된다.

참고문헌

- 김윤중, 김원영, 유일현 (1992) 광역 지질재해분석(산사태)을 위한 GIS 활용. 대한지질공학회지, 2권, p. 131-140.
- 김윤중, 김원영, 유일현 (1993) 지질재해 분석을 위한 GIS 응용연구. 한국GIS학회지, 1권, p. 89-94.
- 김원영 외 (2003) 산사태 예측 및 방지기술 연구. 한국지질자원연구원 보고서 KR-2003 T-03, 339p.
- 박노욱, 지광훈, Chung, C.F., 권병두 (2003) 퍼지 이론을 이용한 GIS 기반 자료우도형 지질자료 통합의 이론과 응용. 자원환경지질, 36권, p. 243-255.
- 박용원, 김감래, 여운광 (1993) 1991년 용인-안성 지역 산사태 연구. 한국지반공학학회지, 9권, p. 103-116.
- 이사로 (2000) 지리정보시스템(GIS)을 이용한 산사태 취약성 분석 기법 개발 및 적용 연구. 연세대학교 박사학위 논문, 163p.
- 토질안전재료위원회 (1990) 사면 안정 공법. 창우출판, 321p.
- 한대석, 이사로, 김경수, 최영섭, 유일현 (1998) GIS를 이용한 한국 동남부지역의 상대적 사면안정성 분류도 작성. 한국GIS학회지, 6권, p. 25-33.
- Bonham-Carter, G.F. (1994) Geographic information systems for geoscientists: modelling with GIS. Pergamon Press, Oxford, 398p.
- Chung, F.C. and Fabbri, A.G. (1998) Three Bayesian prediction models for landslide hazard. Proceeding of International Association for Mathematical Geology (IAMG'98), p. 204-211.
- Chung, F.C. and Fabbri, A.G. (1999) Probability prediction models for landslide hazard mapping. Photogrammetric Engineering & Remote Sensing, v. 65, p. 1389-1399.
- Chung, F.C., Kojima, H. and Fabbri, A.G. (2002) Stability analysis of prediction models for landslide hazard mapping. In Allison, R.J. (ed.) Applied geomorphology: theory and practice. John Wiley & Sons, Inc., London, p. 3-19.
- Duda, R.O., Hart, P.E. and Stork, D.G. (1997) Pattern classification. John Wiley & Sons, Inc., New York, 654p.
- Lee, S. and Min, K. (2001) Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental Geology, v. 40, p. 1095-1113.
- Lee, S., Choi, J. and Min, K. (2002) Landslide susceptibility analysis and verification using the Bayesian probability model. Environmental Geology, v. 43, p. 120-131.
- McLachlan, G.J. (1992) Discriminant analysis and statistical pattern recognition. John Wiley & Sons, Inc., New Jersey, 526p.
- Park, N.-W., Chi, K.-H., Chung, C.F. and Kwon, B.-D.

- (2003) Application of spatial data integration based on the likelihood ratio function and Bayesian rule for landslide hazard mapping. *Journal of the Korean Earth Science Society*, v. 24, p. 428-439.
- Parzen, E. (1962) On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics*, v. 33, p. 1065-1076.
- Pearl, J. (1997) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers, San Mateo, Calif., 552p.
- SpatialModels Inc. (2004) User's guide of spatial prediction modeling system. 108p.
-
- 2004년 10월 12일 원고접수, 2005년 1월 17일 게재승인.