

## 산사태 발생예측을 위한 확률모델

### A Probabilistic Model for Landslide Prediction

채병곤<sup>1)</sup>, Byung-Gon Chae, 김원영<sup>2)</sup>, Won-Young Kim, 조용찬<sup>3)</sup>, Yong-Chan Cho, 송영석<sup>1)</sup>, Young-Suk Song

<sup>1)</sup> 한국지질자원연구원 지질환경재해연구부 선임연구원, Senior Researcher, Geol. & Envir. Hazards Division, Korea Institute of Geoscience & Mineral Resources (KIGAM)

<sup>2)</sup> 한국지질자원연구원 지질환경재해연구부 책임연구원, Principle Researcher, Geol. & Envir. Hazards Division, Korea Institute of Geoscience & Mineral Resources (KIGAM)

<sup>3)</sup> 한국지질자원연구원 지질환경재해연구부 연구원, Researcher, Geol. & Envir. Hazards Division, Korea Institute of Geoscience & Mineral Resources (KIGAM)

**SYNOPSIS** : In this study, a probabilistic prediction model for debris flow occurrence was developed using a logistic regression analysis. The model can be applicable to metamorphic rocks and granite area. In order to develop the prediction model, detailed field survey and laboratory soil tests were conducted both in the northern and the southern Gyeonggi province and in Sangju, Gyeongbuk province, Korea. The six landslide triggering factors were selected by a logistic regression analysis as well as several basic statistical analyses. The six factors consist of two topographic factors and four geological and geotechnical factors. The model assigns a weight value to each selected factor. The verification results reveal that the model has 86.5% of prediction accuracy. Therefore, it is possible to predict landslide occurrence in a probabilistic and quantitative manner.

**Key words** : probabilistic prediction model, debris flow, logistic regression analysis, landslide triggering factors

## 1. 서론

국내에서 발생하는 산사태는 대부분 토석류(debris flow)에 해당하며, 7, 8월에 집중되는 호우에 기인되는데 우리나라의 연 평균 강수량 1,200mm 중의 절반 이상이 7, 8에 집중되고 최근 3~4년간 경기도 일대에서 측정된 강우자료에 의하면 2~3일간에도 최대 600mm 이상이 기록된 바도 있다(김원영 외, 2003). Olivier et al.(1994)는 24시간 동안의 강우량이 연 평균 강수량의 20%를 초과할 경우 대형 산사태가 일어날 수 있다고 보고한 바 있다. 그러나, 집중강우가 산사태의 가장 직접적인 원인이지만 동일 강우 지역 내에서 지질요소의 특성에 따라 산사태 발생빈도와 양상이 달라진다(千木良, 1997; Dikau et al., 1996). 그러므로, 산사태 발생을 예측하기 위해서는 지질요소별 특성을 파악하고 그에 적합한 산사태 예측모델을 설정할 필요가 있다.

이 연구에서는 지질별로 적용가능한 산사태 예측모델을 설정하기 위해 국내에 가장 넓게 분포하는 대표 암석종류 중 경기도를 중심으로 편마암으로 대표되는 변성암류와 경기도 일부지역과 경북 일부지역

에 분포하는 화강암을 대상으로 산사태 예측모델을 개발하였다(김원영 외, 2003, 김원영 외 2004). 또한, 이 지역들은 지난 수년간 산사태 발생이력이 가장 많은 곳으로서 막대한 재산과 인명의 피해가 발생하였다. 산사태 자료는 경기북부와 남부지역을 대상으로 1,700여개의 산사태를 조사하고, 경북 상주 일대에서 99개의 산사태를 추가 조사하였다. 이 자료를 토대로 통계적 기법을 이용하여 지질특성을 고려한 확률론적 산사태 예측모델을 제시하고자 한다.

## 2. 산사태 자료 통계분석

### 2.1 변수설명

표 1은 경기도 및 상주 지역에서 산사태 발생(목표변수)에 영향을 미치는 14개 조사항목(입력변수)을 정리한 것이다. 분석에 사용될 총 14개의 변수 중 4개는 지형관련, 10개는 지질관련 변수로 구성되어있다.

표 1. 변수명과 변수설명(\*:범주형변수를 의미)

목표 변수	*산사태 발생 여부(Target, 1: 발생 0:미발생)	
입력 변수	지형(4)	*암석종류(Lithology), 지형고도(Elevation) 사면방향(Direction), 사면경사(Angle)
	지질(10)	비중(Gravity), 함수율(Moisture), 간극비(VoidRatio), 공극률(Porosity), 습윤도(DegSaturation), 전체밀도(WetDensity), 포화밀도(SatDensity), 건조밀도(DryDensity), *입도분포(USCS), 투수계수(Permeability)

### 2.2 기초통계분석

표 2와 표 3은 경기도 및 상주지역에서 관측한 190 개체(결측치가 제거되지 않은 전체자료)에 대한 정보를 요약한 표이다. 관측된 자료를 살펴보면 Lithology는 거의 'Gneiss'와 'Granite'로 구성되어 있으며 USCS 변수는 'SC'와 'SW-SC'가 전체의 85.71%로 구성되어 있다. 나머지 연속형 변수들에 대한 평균 및 분산은 표 3에 정리하였다.

표 2. 범주형 변수에 대한 기초통계량

Variables		Frequency(N)	Percentage(%)
Lithology	Gneiss	120	63.16
	Granite	66	34.74
	Schist	4	2.11
USCS	CL	1	0.79
	SC	74	58.73
	SC-SM	4	3.17
	SM	6	4.76
	SW	6	4.76
	SW-SC	34	26.98
	SW-SM	1	0.79

표 3. 연속형 변수에 대한 기초통계량

Variables	Frequency(N)	Average	S.D	Min.	Max.
*Elevation	186	325.091	121.919	70.00	655.00
Direction	157	186.636	105.970	1.00	350.00
*Angle	185	30.621	8.561	7.59	48.00
Gravity	92	2.665	0.044	2.56	2.76
Moisture	92	18.313	4.895	8.72	33.47
*VoidRatio	89	1.073	0.311	0.71	3.21
*Porosity	123	77.178	7.465	53.10	94.98
*DegSaturation	89	47.891	13.685	21.15	89.88
*WetDensity	89	1.557	0.151	1.29	1.87
SatDensity	89	1.827	0.091	1.64	2.09
*DryDensity	123	1.333	0.117	1.03	1.58
Permeability	123	0.128	0.650	1.54×10 <sup>-6</sup>	6.34

위의 표 3에서 \*표시가 된 변수는 아래의 유의수준 5%에서의 T-검정 결과 유의한 변수를 나타낸다. 유의한 정도를 나타내는 구체적인 수치는 표 4와 같다.

또한, 각 변수 별로 산사태가 발생된(Target=1) 그룹과 발생하지 않은(Target=0) 그룹간에 평균을 비교한 결과 표 4와 같이 Elevation, Angle, Void ratio, Porosity, Degree of saturation, Wet density, Dry density가 유의수준 5%에서 유의한 변수로 나타났다. 이는 산사태 여부에 따라 각 유의한 변수의 평균 차이가 있다는 것을 의미한다.

표 4. 유의수준 5%하에서 T-검정 결과(괄호속의 값은 유의확률을 나타냄.)

Effective variables	Elevation(0.0321), Angle(<.0001), Void ratio(0.0029), Porosity(0.0089) Deg. saturation(0.0152), Wet density(0.0145), Dry density(0.0021)
Not effective variables	Direction(0.4346), Gravity(0.1741), Moisture(0.4873) Sat. density(0.2321), Permeability(0.5936)

### 3. 로지스틱 회귀분석

산사태 발생여부(목표변수, Target)에 영향을 미치는 입력변수의 영향을 모형화 하기 위해 로지스틱회귀 분석을 실시하였다. (로지스틱 회귀분석은 목표변수가 이진형인 경우에 적합한 회귀분석으로 알려져 있다.) 총 14개의 입력변수가 사용되었으며 조사된 자료 가운데 결측치가 5개 이상인 관측치는 분석에서 제외하였다.

분석에 이용된 자료는 총 89 개체이며, 정보손실을 줄이기 위해 Elevation과 Angle 변수에서 발생된 4개의 결측 자료에 대해서는 해당 그룹의 평균값으로 대체하였다. Lithology 가 Schist인 3개의 값은 그와 유사한 Gneiss로 통합하였다.

조사된 14개변수를 이용하여 적합된 최적의 회귀분석의 결과는 표 5와 같다. 이 결과는 모든 가능한 회귀의 결과를 바탕으로, 전문가의 조언 그리고 전진선택법과 후진제거법을 통한 변수선택 순서 등을 고려하여 최종적으로 선택된 최적모형이다. 이 모형에는 총 6개의 입력변수가 포함되었으며, 그 가운데 지형변수가 3개(Lithology, Elevation, Angle), 지질변수가 3개(Porosity, Dry density, Permeability) 로 나타나고 있다. 특히, T-검정에서도 가장 유의한 것으로 판명된 Angle 변수가 가장 영향이 큰 변수로 선택되었다.

표 5. 로지스틱 회귀분석 결과

Parameter	DF	Estimate	Standard Error	Wald Chi-square	Pr > Chi-square
Intercept	1	-9.3670	7.6137	1.51	0.2186
Angle	1	0.2129	0.0455	21.87	<.0001
Dry density	1	0.7690	3.0486	0.06	0.8009
Elevation	1	-0.0052	0.0036	2.08	0.1491
Lithology(Gneiss)	1	0.4248	0.7630	0.31	0.5777
Permeability	1	0.1777	0.4170	0.18	0.6699
Porosity	1	0.0555	0.0517	1.15	0.2832

위의 결과로부터 적합된 로지스틱 회귀모형은 다음과 같이 표현된다. 산사태가 일어날 확률의 추정치를  $p = \hat{P}$ (산사태가 일어날 확률)라 하자.

$$\text{logit}(p) = -9.3670 + 0.2129 \cdot \text{Angle} + 0.7690 \cdot \text{Dry density} - 0.0052 \cdot \text{Elevation} + 0.4248 \cdot \text{Lithology} + 0.1777 \cdot \text{Permeability} + 0.0555 \cdot \text{Porosity} \quad (1)$$

(주의: Lithology의 값은 Gneiss의 경우는 1이고, Granite의 경우는 0임.)

위 식은 다시 아래와 같이 표현될 수 있다.

$$p = 1 / [1 + \exp\{-\text{logit}(p)\}] \quad (2)$$

$$\text{또는 } p = \exp\{\text{logit}(p)\} / [1 + \exp\{\text{logit}(p)\}] \quad (3)$$

위의 식으로부터 우리는 특정지역의 산사태 발생 가능성을 다음과 같이 예측할 수 있다. 예를 들어, 원자료에서 산사태가 발생한(Target=1) 것으로 조사된 MAP-ID가 AS51-04 인 지역에서의 산사태 발생확률은 위의 모형으로부터 약 87%인 것으로 추정될 수 있다(표 6).

표 6. AS51-04 지역의 관측데이터와 추정확률

MAP-ID	Lithology	Elevation	Angle	Porosity	DryDensity	Permeability	Target	Inference
AS51-04	Granite	204	32.00	76.26	1.210	5.77E-06	1 (Occurred)	0.87

(계산과정)

$$\begin{aligned} \text{logit}(p) &= -9.3670 + 0.2129 \cdot 32.0 + 0.7690 \cdot 1.210 - 0.0052 \cdot 204 \\ &\quad + 0.4248 \cdot 1 + 0.1777 \cdot 5.77\text{E-}06 + 0.0555 \cdot 76.26 = 1.9035 \end{aligned}$$

$$\text{따라서, } p = \exp(1.9035) / [1 + \exp(1.9035)] = 0.870$$

이다.

위에서 적합된 로지스틱 회귀식을 이용하여 각 개체별로 산사태가 발생할 확률이 얼마인지를 알 수 있다. 만약 해당개체로부터 추정된 확률값이 0.5(분류기준값) 보다 크면, 산사태가 일어날 가능성이 더 높으므로, 산사태가 일어나는 것으로 판단하고, 0.5 보다 작은 경우는 산사태가 일어나지 않는 것으로 판정할 수 있다. 즉, 로지스틱 회귀분석을 통해 산출되는 확률의 추정값(이를 사후확률이라고도 함.)으로부터 다음과 같은 정오분류표(confusion matrix)를 얻을 수 있다(표 7).

표 7. 로지스틱 판별분석에 대한 정오분류표

Inference result Observed results	Occurred (Target=1)	Not occurred (Target=0)	Total
Occurred (Target=1)	47	5	52
Not occurred (Target=0)	7	30	37
Total	54	35	89

표 7은 추정된 회귀모형이 자료를 얼마나 정확하게 분류해 내는가를 나타내는 것으로, 모형에 대한 성능을 평가하는 잣대가 될 수 있다. 즉, 위의 분류표로부터 대각선에 나타난 수치(47, 30)은 추정된 회귀모형으로부터 제대로 분류된 경우를 말하며, 반대방향의 자료(5, 7)은 오분류(missclassification)한 경우의 정보를 나타낸다. 따라서, 전체적으로 위 모형을 이용하면 86.5%의 예측률을 가짐을 알 수 있고, 이는 산사태 예측도가 매우 높은 정확성을 가짐을 표시한다.

#### 4. 토의 및 결론

통계적 측면의 검증 외에 실제 야외조사 자료와 산사태 발생가능성 예측도를 비교해 보았다(그림. 1). 예측도가 작성된 두 지역에서 실제 산사태가 발생한 지점이 대부분 75%이상의 발생가능성을 가진 영역에 접시 되었다(채병곤 외, 2004). 75% 이상의 발생확률을 가진 지점은 실제 지반상태가 매우 불안정한 곳으로서, 산사태 발생 인자간의 미세한 변화가 발생하거나, 약간의 외부 영향만 가해지더라도 쉽게 산사태가 일어날 가능성이 있는 곳이다. 현장에서 관찰한 모든 산사태 시작부의 위치가 예측도상의 발생가능성이 높은 불안정영역과 일치함을 통해 이 연구에서 선정한 산사태 발생인자와 그 인자별 가중치의 부여가 정확하였음을 알 수 있다. 이러한 정확성은 현장에서 획득한 산사태 미발생지점의 위치가 발생가능성이 낮은 지점에 표시됨을 통해서도 다시 한번 입증된다.

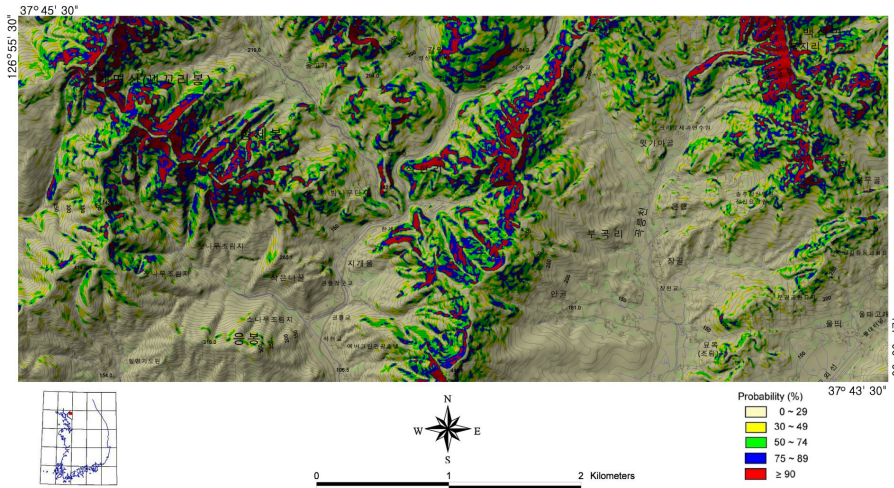


그림 1. 로지스틱 회귀모델을 이용해 작성한 산사태 예측도

통계기법을 산사태에 적용한 사례는 극히 드물다. 이번 연구에서는 산사태 발생인자의 체계적이고 객관적인 선정과 이에 대한 가중치 설정을 위해 로지스틱 회귀분석을 이용하였다. 이를 통해 산사태에 영향을 미치는 인자들을 추출하였고, 이 인자들을 포함한 발생가능성 판별식도 적합함을 알아냈다. 따라서, 위에서 구한 회귀모형을 이용하여 다른 지역의 산사태 발생여부를 예측하는 작업에 신뢰성을 가지고 접근해 볼 수 있는 계기를 마련하게 되었고, 이 회귀모형을 통한 예측기술은 변성암과 화강암 분포지의 산사태 발생가능성을 예측하는데 무리 없이 적용할 수 있다.

## 감사의 글

본 연구는 한국지질자원연구원에서 지원하는 “산사태 위험도 산정시스템 및 피해저감기술 개발”에 의하여 실시된 연구 결과임을 밝히며, 이에 감사의 뜻을 전하는 바이다.

## 참고문헌

1. 김원영, 채병곤, 김경수, 조용찬, 최영섭, 이춘오, 이철우, 김구영 (2003) “산사태 예측 및 방지기술 연구”, 과학기술부, KR-03-(T)-03.
2. 김원영, 채병곤, 조용찬, 김경수, 이춘오, 송영석 (2004) “산사태 위험도 산정시스템 및 피해저감기술 개발”, 국무총리, KR-04(연차)-13-3.
3. 채병곤, 김원영, 조용찬, 김경수, 이춘오, 최영섭 (2004) “토석류 산사태 예측을 위한 로지스틱 회귀모형 개발”, 지질공학, 제14권, 제2호, pp.211~222.
4. Dikau R., Brunsten, D., Schrott, L. and Ibsen, M.-L. (1996) “Landslide Recognition”, John Wiley and Sons, New York, pp.149~188.
5. Olivier, M. Bell, F. G. and Jemy, C. A. (1994) “The effect of rainfall on slope failure, with examples from the Greater Durban area”, *Proceedings 7th intern. Cong. IAEG*, Vol. 3, pp.1629-1636.
6. 千木良雅弘 (1997) "風化と崩壊", 近未来社, 名古屋, pp.49-87.