

토석류 산사태 예측을 위한 로지스틱 회귀모형 개발

채병곤*·김원영·조용찬·김경수·이춘오·최영섭

한국지질자원연구원 지질환경재해연구부

Development of a Logistic Regression Model for Probabilistic Prediction of Debris Flow

Byung-Gon Chae* · Won-Young Kim · Yong-Chan Cho
Kyeong-Su Kim · Choon-Oh Lee · Young-Sup Choi

Geological & Environ. Hazards Div., Korea Inst. Geoscience and Mineral Res.

이 연구는 자연사면에서 발생하는 토석류(debris flow) 산사태의 확률론적 예측을 위해 로지스틱 회귀분석(logistic regression analysis)을 이용하여 변성암 및 화강암 분포지에 적용할 수 있는 예측모형을 개발한 것이다. 산사태 예측모형을 개발하기 위해 경기 남·북부지역과 경북 상주지역에서 발생한 산사태 자료를 현장조사와 실내토질시험을 통해 직접 획득·분석하였다. 산사태 발생에 영향을 미치는 인자는 기초 통계분석은 물론 로지스틱 회귀분석을 실시하여 최종적으로 7개 영향인자를 선정하였다. 이들 7개 인자는 지형요소 2개와 지질 및 토질특성 요소 5개로 구성되어 있고, 각 인자별 가중치를 부여한 점이 큰 특징이다. 개발된 모델은 신뢰성 검증을 수행한 결과 90.74%의 예측율을 확보한 것으로 나타났다. 이 모델을 이용하여 산사태 발생가능성을 확률적·정량적으로 예측할 수 있게 되었다.

주요어: 토석류, 확률론적 예측, 로지스틱 회귀분석, 산사태 예측모형

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 seven landslide triggering factors were selected by a logistic regression analysis as well as several basic statistical analyses. The seven factors consist of two topographic factors and five geological and geotechnical factors. The model assigns a weight value to each selected factor. The verification results reveal that the model has 90.74% of prediction accuracy. Therefore, it is possible to predict landslide occurrence in a probabilistic and quantitative manner.

Key words : debris flow, probabilistic prediction, logistic regression, landslide prediction model

* Corresponding author : bgchae@kigam.re.kr

서론

산사태는 형태나 생성 메커니즘에 의하여 여러 종류로 분류된다. 이를 형태(type)에 따라서 분류할 때 대체로 낙반(fall), 전도(topple), 슬라이드(slide), 측면 퍼짐(lateral spread), 유동(flow)으로 분류하고(UNESCO Working Party, 1990, EPOCH, 1993), 파괴면의 모양(form)에 의하여 분류할 때는 평면(planar), 쐐기(wedge), 계단(stepped), 원호(circular), 준원호(semi-circular) 등으로 분류할 수 있다. 또한 이들의 생성이 복합적인 작용에 의한 것인가 아닌가에 따라 단일(single) 또는 복합(multiple)이라는 용어를 사용하기도 한다(Hutchinson, 1988).

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 동일강우 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 최근 수년간 수행해 온 산사태 예측연구를 통해 설정한 지질별 정량적 산사태 예측모델을 제시하고자 한다.

연구지역

지질별로 적용가능한 산사태 예측모델을 설정하기 위해 국내에 가장 넓게 분포하는 대표암석종류 중 경기도를 중심으로 편마암으로 대표되는 변성암류, 경기도 일부지역과 경북 일부지역에 분포하는 화강암을 대상으로 산사태 예측모델을 개발하고자 하였다. 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 상주일대에

서 99개의 산사태를 추가 조사하였다.

발생인자 선정 및 가설설정

변수선정 및 가설설정

통계적 분석 기법을 사용하여 사태 발생을 예측하는데 있어 가능성과 타당성을 검토해 보기 위해 경기도 용인-안성 지역의 77개 지점에서 수집된 13개 변수(건조밀도, 암질, 입도분포(gravel, sand, silt), 액성한계, 습윤도, 소성지수, 소성한계, 투수계수, 공극률, 포화도, 간극비, 습윤밀도, 포화밀도)를 사용하여 사전 분석을 실시한 결과, 통계적 분석 방법이 그 동안 지질학적인 측면에서 논의되어 왔던 가설을 만족하면서 유의한 결론을 도출해 낼 수 있다는 가능성을 확인하고(김원영 외, 2000) 다음과 같은 분석을 실시하였다.

그 동안의 사전 분석 및 타당성 검토를 통해, 산사태 발생에 영향을 주는 요인들로서 토질자료로는 암질, 공극률, 건조밀도, 입도분포, 투수계수를, 지형자료로는 고도, 사면방향, 사면경사, 사면 상부경사를 각각 선정하였다. 직관적으로 산사태 발생에 직접적인 영향을 주는 요인은 강우량으로 생각된다. 그러나, 이 연구에서는 강우의 영향을 제외하고 지질요인들 중 산사태에 영향을 미치는 요인을 파악하기 위해 경기지역 내 강우량이 동일한 지역을 대상으로 정밀 분석을 실시하였다. 또한, 화강암 분포지역에서의 산사태 예측 모델 개발을 위해 경기지역 외 경북 보은~상주지역을 추가 대상지역으로 선정하였다.

산사태에 영향을 미치는 변수들의 상호연관성과 이 변수들이 산사태 발생에 직접적인 관련성이 있는지를 통계적 접근 방법을 통해 분석해 보고, 더 나아가 산사태 발생을 확률적으로 예측하는데 있어 다음과 같은 연구가설을 세웠다.

- [가설 I] 토질은 산사태 발생에 유의적인 영향을 미친다.
- [I-1] 암질(Lithology)은 산사태 발생에 유의적인 영향을 미친다.
 - [I-2] 공극률(Porosity)은 산사태 발생에 유의적인 영향을 미친다.
 - [I-3] 건조밀도(Dry density)는 산사태 발생에 유의적인 영향을 미친다.
 - [I-4] 입도분포(USCS)는 산사태 발생에 유의적인 영향을 미친다.

- [I -5] 투수계수(Permeability)는 산사태 발생에 유의적인 영향을 미친다.
- [가설 II] 지형은 산사태 발생에 유의적인 영향을 미친다.
- [II -1] 고도(Elevation)는 산사태 발생에 유의적인 영향을 미친다.
- [II -2] 사면방향(Slope dip direction)은 산사태 발생에 유의적인 영향을 미친다.
- [II -3] 사면경사(Slope angle)는 산사태 발생에 유의적인 영향을 미친다.
- [II -4] 사면 상부경사(Slope top angle)는 산사태 발생에 유의적인 영향을 미친다.

자료 특성 및 분석방법

산사태 발생예측을 위한 자료는 경기북부 지역의 총 표본수 116개로 산사태가 발생한 지점 79개소, 산사태 미발생 지점 37개소를 대상으로 분석하였다. 또한, 화강암 지역에 동시에 적용하기 위한 모델개발을 위해 경기남부 자료 외에 경북 상주지역을 대상으로 산사태가 발생한 지점 29개, 산사태 미발생 지점 29개 등 총 표본수 58개를 부가하여 분석을 실시하였다. 그리고, 위에서 보는 바와 같이 산사태에 영향을 미치는 변수로 총 9개의 변수가 선정되었다. 9개의 변수 중 입도분포와 암질은 수치 데이터가 아닌 명목형 데이터이며, 그 외 7개의 변수는 모두 수치 데이터이다. 먼저 변수들의 상호 연관성 및 구조적 특징을 파악하기 위하여 '주성분 분석'을 실시하고, 주성분 분석을 통하여 몇 개의 주성분을 판별변수로 이용할 수 있을지를 점검해 보았다. 이 분석의 궁극적인 목적은 통계적으로 산사태에 유의한 영향을 미치는 변수들을 찾아내고, 이 변수들을 토대로 산사태 발생 여부를 예측하는데 있다. 그러한 예측을 위한 통계기법의 하나인 '로지스틱 회귀분석'을 통하여 단계적으로 유의한 변수들을 선택하고 선택된 변수들로 이루어진 하나의 판별함수를 구해 보았다. 분석용 통계 패키지는 SPSS 9.0과 SAS 6.12를 사용하였다.

주성분 분석을 이용한 자료분석

주성분분석이란 여러 변수가 서로 상관관계를 가지고 있을 때 이들을 설명할 수 있는 구조를 찾아내는 방법이다. 이 분석에서는 서로 상관관계를 갖고 있는 변

수들을 이용해서 상관관계를 갖지 않는 새로운 변수를 찾아내는 것이 목적이다. 이러한 변수들을 주성분이라 한다. 다시 말해, 주성분 분석은 다차원의 데이터를 정보의 손실을 가능한 적게 하면서 2차원 혹은 3차원의 데이터로 축약하는 기법이다. 주성분 분석은 데이터 유형이 수치 데이터인 경우에 한하여 분석이 가능하므로 여기서는 입도분포와 암질을 제외한 7개의 변수만으로 분석하였다.

단순통계를 통한 각 7개 변수의 평균과 표준편차를 구하였으며(Table 1), 이 변수들간의 상관관계는 Table 2와 같다. 주성분 분석결과 Table 3과 같이 3개의 주성분이 가장 적합한 것으로 나타났다. 즉, 우리는 7개의 변수를 3개의 변수로 줄여서 분석하여도 정보의 손실이 적으면서도 차원의 수는 줄여 간단하게 분석을 할 수 있다.

3개의 주성분은 다음과 같이 얻어진다.

$$C_1 = -0.082u_1 + 0.115u_2 + 0.185u_3 + 0.457u_4 - 0.135u_5 + 0.604u_6 + 0.596u_7 \quad (1)$$

$$C_2 = 0.685u_1 - 0.642u_2 + 0.239u_3 - 0.091u_4 - 0.191u_5 + 0.049u_6 + 0.121u_7 \quad (2)$$

$$C_3 = 0.125u_1 - 0.292u_2 - 0.658u_3 + 0.202u_4 + 0.625u_5 + 0.126u_6 + 0.136u_7 \quad (3)$$

단, u_1 =Porosity, u_2 =Dry density,

u_3 =Permeability, u_4 =Elevation

u_5 =Slope dip direction, u_6 =Slope dip angle,

u_7 =Slope top angle

누적 기여율을 통해 3개의 주성분만으로도 원래의 모든 변수가 가지고 있는 정보의 71.79%를 설명할 수 있다는 것을 알 수 있다(Table 4). 그러나, 주성분 분석을 통하여 몇 개의 주성분을 판별변수로 이용하는 방법은 해석이 어려워지고, 또한 서로 상관관계가 아주 높은 변수도 사면경사와 사면 상부경사(상관계수 = 0.877)뿐이므로 산사태 예측에 있어 그다지 효율적인 분석 방법은 아니라 생각된다.

Table 1. Result of simple statistics for the seven variables

		Average	Standard Deviation
Soil Property Data	Porosity	76.686	6.828
	Dry density	1.307	0.142
	Permeability	0.00018	0.00074
Geomorphologic Data	Elevation	244.505	0.163
	Slope dip direction	201.794	100.998
	Slope dip angle	29.262	10.230
	Slope top angle	29.533	10.530

Table 2. Result of Pearson's correlation coefficients for the seven variables

	Porosity	Dry density	Permeability	Elevation	Slope dip direction	Dip angle	Slope top angle
Porosity	1.0000	-.5370	0.1563	-.1222	-.0150	-.0475	0.0096
Dry density	-.5370	1.0000	0.0623	0.1038	0.0637	0.0927	0.0034
Permeability	0.1563	0.0623	1.0000	0.0063	-.2074	0.1828	0.1849
Elevation	-.1222	0.1038	0.0063	1.0000	-.0822	0.4850	0.4702
Slope dip direction	-.0150	0.0637	-.2074	-.0822	1.0000	-.0779	-.1088
Dip angle	-.0475	0.0927	0.1828	0.4850	-.0779	1.0000	0.8766
Slope top angle	0.0096	0.0034	0.1849	0.4702	-.1088	0.8766	1.0000

Table 3. Result of principal component analysis

	PRIN 1	PRIN 2	PRIN 3
Porosity	-.082344	0.685110	0.124562
Dry Density	0.114745	-.641661	-.291621
Permeability	0.184691	0.238792	-.658334
Slope dip direction	-.134739	-.191211	0.625112
Slope dip angle	0.603596	0.049179	0.126397
Slope top angle	0.595808	0.120898	0.135952
Elevation	0.456587	-.090980	0.202022

Table 4. Cumulative contribution ratio

	Original value	Difference	Ratio	Cumulative contribution ratio
PRIN 1	2.32962	0.753532	0.332803	0.33280
PRIN 2	1.57609	0.456336	0.225156	0.55796
PRIN 3	1.11976	0.252853	0.159965	0.71792
PRIN 4	0.86690	0.284346	0.123843	0.84177
PRIN 5	0.58256	0.175954	0.083222	0.92499
PRIN 6	0.40660	0.288140	0.058086	0.98308
PRIN 7	0.11846		0.016923	1.00000

로지스틱 회귀분석 (Logistic Regression)

회귀분석은 산사태와 관련 있는 다른 요인들을 이용하여 향후 발생가능한 산사태를 예측하려는 시도에 적용 가능한 통계분석 방법이다. 예측하려는 변수 즉, 산사태의 발생가능성을 '종속변수'라 하고 산사태를 예측하기 위해 사용하는 다른 요인을 '독립변수'라 한다. 산사태 발생여부를 예측하기 위해서는 종속변수가 '산사태 발생함', '산사태 발생 안함'의 단 두 개뿐이므로 일반적인 회귀분석을 그대로 적용할 수는 없고 이 때 사용하는 분석방법이 바로 '로지스틱 회귀분석'이다. 즉, 로지스틱 회귀분석은 단지 두 개의 값만을 가지는 종속변수와 다양한 값을 갖는 독립변수들 사이의 관련성을 분석하는 통계기법이다(성용현, 2001). 또한, 로지스틱 회귀분석은 산사태가 발생할 경우와 발생하지 않을 경우를 예측하기보다는 그 지역에 산사태가 발생할 확률을 계산해 준다는 점에서 산사태 분석에 있어 가장 이상적인 통계 분석방법이라 할 수 있다.

통계적 이론

일반적으로 회귀분석에서의 모형은 주어진 독립변수 하에서 종속변수의 평균이 독립변수에 대한 선형식으로 표현된다. 즉, 아래와 같은 식으로 표현되는 것

이다.

$$E(y | x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (4)$$

그러나, 예를 들어 산사태 자료의 경우, 산사태 발생 유무로 표현되는 이분형 종속변수인 경우에 $E(y | x)$ 는 독립변수가 x 로 주어진 경우에 산사태가 발생할 확률을 의미하게 된다.

일반 회귀모형에서는 x 의 범위가 $-\infty$ 에서 $+\infty$ 의 값을 가지면 y 의 범위 역시 $-\infty$ 에서 $+\infty$ 까지의 값을 가지나 위에서 $E(y | x)$ 는 확률이므로 0에서 1 사이의 값만을 가질 수 있으며, $E(y | x)$ 는 독립변수의 선형식으로 표현되기보다는 S 형태를 따르는 곡선형태에 가깝다. 이러한 반응함수의 제약성 $0 \leq E(y | x) \leq 1$ 을 충족시켜주는 함수는 몇 가지 종류가 있으나, 이들 함수들 중에서 사용이 용이하다는 점에서 다음의 로지스틱 함수를 사용하게 된다.

$$E(y | x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (5)$$

위의 로지스틱 함수는 β_0 와 β_1 에 대해서 비선형 함수이나 이를 선형으로 변환시킬 수 있다. 기대반응 $E(y | x)$ 는 확률을 의미하므로

$$E(y | x) = p \quad (6)$$

로 놓자. 그러면, 다음의 변환

$$p' = \ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{E(y|x)}{1-E(y|x)}\right) \quad (7)$$

을 이용하여 식 (5)를 식 (7)에 대입시키면

$$p' = \beta_0 + \beta_1 x \quad (8)$$

으로 뒤를 알 수 있다. 이와 같은 변환을 로지스틱 변환(logistic transformation)이라 하고, p' 즉,

$$\ln\left(\frac{p}{1-p}\right) \text{를 로짓(logit)이라 부른다.}$$

위에서 언급하였던 바와 같이 산상태 자료는 종속 변수가 이분형 자료이고 독립변수가 연속형 자료이기 때문에 일반 회귀모형을 그대로 적용하기가 매우 어려웠으나 이 자료를 가지고 로지스틱 변환시키면 일반적인 회귀모형의 형태를 지니게 되어 분석이 용이하게 된다.

로지스틱 회귀계수의 추정과 검정

로지스틱 회귀계수는 다른 선형회귀계수와 마찬가지로 종속변수와 독립변수들 사이의 관계를 설명하고 주어진 독립변수의 수준에서 종속변수를 예측하는데 사용되어 진다. 그러나, 회귀계수의 추정방법에 차이가 있다. 선형분석에서는 잔차의 제곱합(sum of residual)을 최소화 하지만, 로지스틱 분석은 우도(likelihood) 즉, 사건발생 가능성을 크게 하는데 있다 (성용현, 2001).

로지스틱 회귀계수를 추정하는 방법은 독립변수의 수준에서 반복적인 종속변수 관측여부에 따라 달라지는데, 각 독립변수의 수준에서 비교적 많은 종속변수의 반복적인 관측이 있으면 가중 최소자승법(weighted least squares method)을 사용하고, 반복적인 관찰이 없거나 아주 작은 경우에는 최대우도추정법(maximum likelihood estimation method)을 사용한다.

■가중 최소자승법(weighted least squares method)

가중최소자승 추정방법은 주어진 독립변수의 수준

에서 반복적인 종속변수의 관측자료가 주어진 경우에 사용된다. 예를 들어, 독립변수가 하나인 경우 관찰된 x 수준이 k 개 있다고 가정하자. 각 수준 x_i ($i = 1, 2, \dots, k$)에서 종속변수 y 에 대한 반복적인 관찰회수를 n_i 라 하자. 이 때 독립변수 x_i 수준에서 y 값이 1인 회수를 r_i 라 하였을 때 x_i 에서 y 값이 1을 취할 표본비율은,

$$p_i = \frac{r_i}{n_i} \quad (9)$$

가 된다. 따라서, 가중 최소자승추정 방법에서는 표

본비율 p_i 를 로짓 변환시킨 $\ln\left(\frac{p_i}{1-p_i}\right)$ 를 종속 변수로 사용한다. 로짓변환은 비선형함수를 선형함수로 변환할 수 있으나 종속변수의 분산이 일정하지 않기 때문에 가중치

$$w_i = n_i p_i \frac{p_i}{1-p_i} \quad (10)$$

를 사용하여 분석을 하게 된다. 표본비율을 사용한 로짓 반응함수는 다음과 같다.

$$p_i' = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} \quad (11)$$

■ 최대우도 추정법(maximum likelihood estimation method)

독립변수의 각 수준에서 y 의 반복적인 관측이 아주 적거나 없으면 표본비율을 사용할 수 없기 때문에 독립변수의 각 수준에서 하나의 y 값에 대하여 최대우도 추정법을 사용하여 로지스틱 반응함수를 추정한다. 일단 최대우도 추정법에 의하여 회귀계수가 추정되면 로지스틱 회귀모형이 자료에 대하여 어느 정도 설명력이 있는지를 검정한다. 로지스틱 회귀모형에서는 다중회귀모형에서 사용한 F-검정과 유사한 우도값 검정(likelihood value test)을 실시한다. 그 순서는 다음과 같다.

가설설정

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

H_1 : 적어도 하나는 0이 아니다.

·우도비 검정통계량 산출

일반적으로 추정된 모형의 적합성은 우도값 검정에 의해서 판단된다. 우도값은 '로그 -2배' 또는 '-2Log Likelihood'(또는 -2LL)라고 한다. -2LL는 모형이 자료에 얼마나 적합한지에 대한 정도를 나타낸다. 단계적 선택법에서 -2LL의 변화량은 모형에서 제거된 항의 계수가 0이라는 가설을 검정한다. 우도값의 식은 다음과 같다.

$$\Lambda = -2 \ln \frac{L_0}{L} = -2 \ln L_0 + 2 \ln L \tag{12}$$

여기서 L 은 k 개의 독립변수들의 정보를 모두 이용한 우도를 나타내며, L_0 는 k 개의 독립변수들이 종속변수의 변화에 전혀 영향을 미치지 못한다고 가

정했을 때의 우도를 나타낸다. 따라서, 모형에 포함된

독립변수들이 중요한 변수가 아니라면 우도비 $\frac{L_0}{L}$

는 거의 같아져서, 우도비 대수함수인 검정통계량 Λ 의 값이 0에 가까운 작은 값을 가지게 된다. 이러한 경우에 우리들은 모형이 적합하지 못하다고 결론을 내릴 수 있다. 반면에 중요한 독립변수가 포함되어 있을 때에는 검정통계량 Λ 의 값은 커지게 된다. 검정통계량 Λ 의 표본분포는 귀무가설이 참일 때, 자유도 $df = k$ 인 χ^2 분포를 따른다.

· 기각치 설정 및 의사결정

유의수준 α 에서 기각치 $\chi^2_{\alpha}(df = k)$ 과 검정통계량의 값을 비교하여 귀무가설 채택 여부를 결정한다.

$\Lambda \leq \chi^2$ 이면 H_0 을 채택한다.

Table 5. Result of logistic regression analysis using stepwise method

Variables	Regression Coefficient	Wald statistics	Degree of freedom	P-value
LITHOLOGY		7.1616	2	.0443
LITHOLOGY(1)	1.4068	3.6699	1	.0266
LITHOLOGY(2)	0.1135	.6913	1	.9031
POROSITY	0.0823	.3739	1	.0923
DRY DENSITY	6.6785	3.8878	1	.0140
USCS		5.9514	4	.5233
USCS(1)	-10.3138	.0095	1	.6770
USCS(2)	-10.0153	.0009	1	.6858
USCS(3)	-11.5952	.0048	1	.6398
USCS(4)	-11.5204	.0111	1	.6421
PERMEABILITY	-1233.79	3.9825	1	.0007
ELEVATION	0.0087	4.4712	1	.0819
SLOPE ANGLE	0.1726	11.7587	1	.0000
Constant	-10.7014	.1009	1	.6765

Variable	Lithology	Porosity	DryDensity	USCS	Permeability	Elevation	Slope Angle
44th data	Gn	83.22	1.479	SM	0.0000364	501	22

$\Lambda \geq \chi^2$ 이면 H_0 을 기각한다.

그리고 추정된 계수의 통계적 유의성 판단은 Wald 통계량으로 한다.

분석결과

효과적인 자료처리와 유의한 변수선택을 위해 단계적 변수 선택법(Stepwise Method)을 이용한 로지스틱 회귀분석을 실시하였다. 단계적 변수 선택법은 유의한 변수를 선정하여 가장 최적의 회귀모형을 추정 해주는 방법으로서, 이 방법을 통해 선택된 산사태에 통계적으로 유의하게 영향을 미치는 변수를 살펴보면 토질자료로는 암질, 건조밀도, 입도분포, 투수계수가 있고 지형자료로는 고도, 사면경사가 선택되었다. 따라서 이 6개의 변수를 이용하여 산사태 예측을 위한 회귀모형을 추정해 볼 수 있다. 한편, 공극률은 입도분포 및 투수계수와 관련성이 높은 항목이므로 공극률 또한 산사태 발생에 중요한 영향을 미치는 변수라고 판단되어 이를 포함하여 총 7개의 변수를 이용하여 로지스틱 회귀모형을 추정해 보았다(Table 5).

$$\begin{aligned} \text{LOGIT} = & 1.4068 \times \text{Lithology 1} + 0.1135 \times \text{Lithology 2} \\ & + 0.0823 \times \text{Porosity} + 6.6785 \times \text{Dry Density} \\ & - 10.3138 \times \text{USCS 1} - 10.0153 \times \text{USCS 2} \\ & - 11.5952 \times \text{USCS 3} - 11.5204 \times \text{USCS 4} \\ & - 1233.79 \times \text{Permeability} + 0.0087 \times \text{Elevation} \\ & + 0.1726 \times \text{Slope angle} - 10.7014 \end{aligned} \quad (13)$$

여기서 Lithology 1: granite, Lithology 2: gneiss
USCS 1: SC, USCS 2: SM, USCS 3: SC-SM, USCS4: CL이다.
식 (13)에서 구한 LOGIT 값을 토대로 한 산사태 발생 확률은 다음과 같다.

$$\text{산사태 발생 확률} = \frac{\text{EXP}(\text{LOGIT})}{(1 + \text{EXP}(\text{LOGIT}))}$$

예를 들어, 데이터의 44번째 지역에서 산사태가 일어날 확률은 다음과 같이 추정된다.

$$\begin{aligned} \text{LOGIT} = & 1.4068 \times 1 + 0.1135 \times 0 + 0.0823 \times 83.22 \\ & + 6.6785 \times 1.479 - 0.3138 \times 1 - 10.0153 \times 0 \\ & - 11.5952 \times 0 - 11.5204 \times 0 - 1233.79 \\ & \times 0.0000364 + 0.0087 \times 501 + 0.1726 \times 22 \\ & - 10.7014 \\ = & 5.2290 \end{aligned}$$

$$\begin{aligned} \text{산사태 발생 확률} = & \frac{\text{EXP}(\text{LOGIT})}{(1 + \text{EXP}(\text{LOGIT}))} \\ = & \frac{\text{EXP}(5.2290)}{(1 + \text{EXP}(5.2290))} \\ = & 0.9947 \end{aligned}$$

산사태 발생하지 않을 확률 = 1 - 0.9947 = 0.0053

따라서, 44번째 지역에서 산사태가 일어날 확률은 99.47%라고 할 수 있다.

산사태 발생예측모델의 신뢰성

발생원인별 가중치의 적합성 검증

산사태 발생원인의 가중치를 적절히 설정하였는지를 검증하는 것은 필수적인 과정이다. 즉, 설정한 예측 회귀모형이 관측한 데이터를 얼마나 잘 분류했는지 결과에 대한 정확성을 살펴보아야 한다. 이를 설명해주는 것이 적합도 통계량(goodness-of-fit statistic)이며 아래와 같다(Table 6).

위에서 7개의 변수와 한 개의 상수항을 포함한 모델의 카이스퀘어 값은 83.150이고 유의수준도 0.0000이므로 위 모형은 통계적인 의미가 있다고 할 수 있다. 즉, 산사태와 고도, 사면경사, 공극률, 건조밀도, 투수계수, 입도분포, 암질과의 관계를 로지스틱 회귀모형으로 분석할 때 적합한 것으로 볼 수 있다.

Table 6. Goodness-of-fit statistics of the proposed logistic model

	Chi-square	Degree of freedom	P-value
Model	83.150	11	.0000
Block	83.150	11	.0000
Step	83.150	11	.0000

추정된 회귀모형이 얼마나 적합한가를 평가하는 또 한 가지 방법은 예측 및 관측치의 분석결과를 비교하는 것이다. 전체 통계처리 자료 중, 산사태가 일어나지 않은 25개 지역은 모형에 의해 산사태가 일어나지 않을 것이라고 옳게 예측된 것을 알 수 있으며, 마찬가지로 산사태가 일어난 65개 지역은 산사태가 일어날 것이라고 옳게 예측된 결과를 보여주고 있다. 표의 역대 각 원소는 얼마나 많은 수의 지역이 잘못 분석되었는가를 말해 준다. 여기에서 모두 9개 지역이 잘못 예측되었다. 따라서, 전체적으로 위 모형을 이용하면 90.74%의 예측률을 가짐을 알 수 있고, 이는 산사태 예측도가 매우 높은 정확성을 가짐을 표시한다.

		예 측		판별력
		발생안함	발생함	
관찰	발생안함	25	4	85.71%
	발생함	5	65	93.15%
전 체				90.74%

통계적 측면의 검증 외에 실제 야외조사 자료와 산사태 발생가능성 예측도를 비교해 보았다(Figs. 1 & 2). 예측도가 작성된 두 지역에서 실제 산사태가 발생한 지점이 대부분 75%이상의 발생가능성을 가진 영역에 점서 되었다. 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 선정된 산사태 발생인자와 그 인자별 가중치의 부여가 정확하였음을 알 수 있다. 이러한 정확성은 현장에서 획득한 산사태 미발생 지점의 위치가 발생가능성이 낮은 지점에 표시됨을 통해서도 다시 한번 입증된다.

적용가능성 검토

통계기법을 산사태에 적용한 사례는 극히 드물다. 이번 연구에서는 산사태 발생인자의 체계적이고 객관

적인 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 무리없이 적용할 수 있다. 다시 말해, 이 기술은 변성암 및 화강암 분포지의 지형, 지질 그리고 토질특성을 모두 포함할 수 있는 대표성을 가지는 것으로서, 기존 연구결과보다 훨씬 정확한 산사태 발생가능성을 예측할 수 있다.

그러나, 이번 연구에서 개발한 산사태 예측기술은 주로 변성암이 분포하는 지점을 대상으로 산사태 가능성을 예측하는 것으로서, 지질조건과 그에 따른 토질특성이 달라질 경우 예측기법을 다소 수정할 필요성이 있다. 또한, 강우조건과 이에 수반한 토양 내 간극수압(pore pressure) 변화요인이 산사태에 가장 큰 영향을 미친다는 그 동안의 연구결과를 감안해, 향후에는 강우 또는 간극수압 등의 변수를 추가하여 예측기법을 수정할 필요가 있다(채병곤 외, 2003). 이렇게 되면 지역에 의존하지 않는 타당하고 안정적인 변수를 포함한 모형이 개발되므로 더욱 정확한 산사태 예측기술을 확보하게 된다.

결 론

이 연구에서는 산사태 특성을 구명하고 발생가능성을 예측하기 위해 기존의 산사태 관련 연구에서 범해왔던 오류를 배제하고, 통계적 기법을 이용해 현장조사 결과 및 실내시험에서 획득한 자료와 인자를 대상으로 각 인자별 상호관련성을 검토하였다. 상호관련성이 높은 인자들을 별도로 추출해서 이들이 산사태 발생에 기여하는 상대가중치를 결정하고, 이를 토대로 실제 산사태가 발생한 지역에 적용하여 통계적 측면에서의 타당성을 검증하였다. 이 과정을 통해 산사태 발생에 큰 영향을 미치는 인자들을 밝혀내고, 이 인자들의 정확한 가중치를 설정하였다.

로지스틱 회귀분석 중 모든 변수를 이용한 분석법을 적용한 결과, 유의수준 $\alpha = 0.05$ 에서 각 추정된 회귀계수들 중 통계적으로 유의하지 않은 변수들이 있어, 단계적 변수 선택법을 통해 산사태에 영향을 미치는

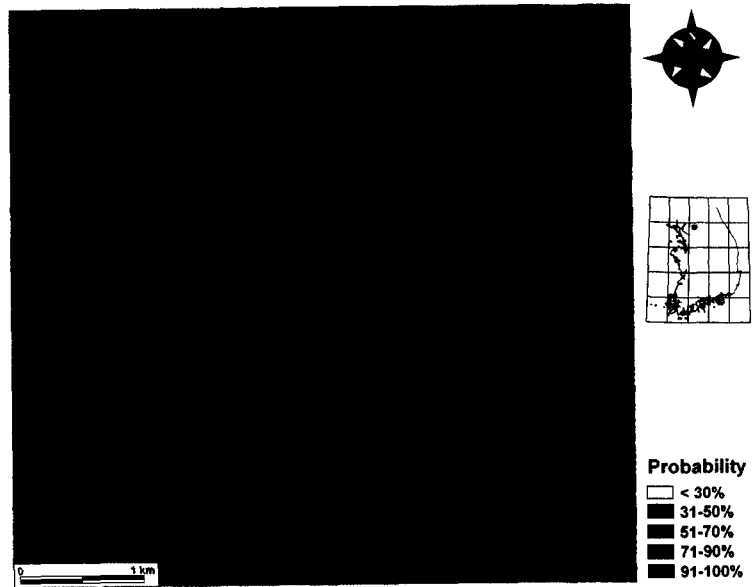


Fig. 1. Landslide probability map of Jinjeob, Gyeonggi province.

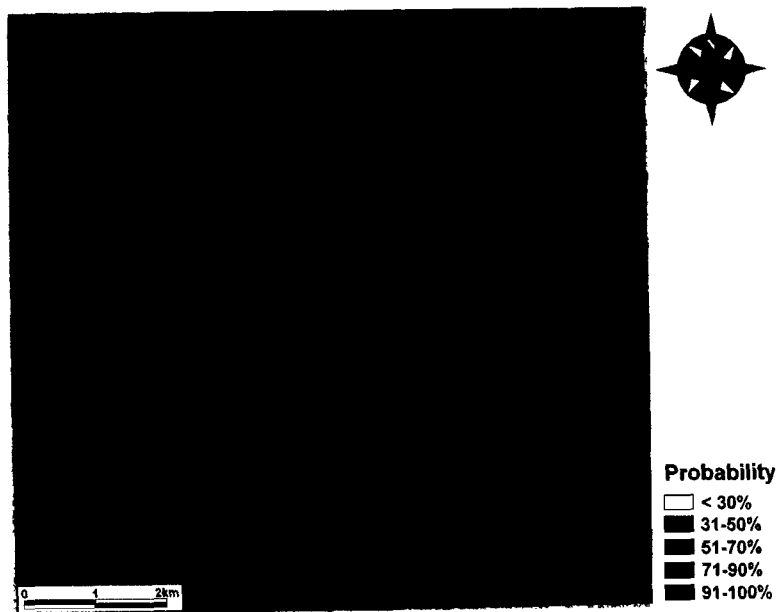


Fig. 2. Landslide probability map of Sangju, Gyeongbuk province.

유의한 변수만 선정하여 회귀모형을 추정하였다. 이 방법을 통하여 산사태에 유의하게 영향을 미치는 것으로 선택된 통계적 변수는 토질자료로는 암질, 건조 밀도, 입도분포 및 투수계수, 그리고 지형자료로는 고도와 경사가 선택되었다. 따라서, 이 6개의 변수를 이용하여 산사태 예측을 위한 회귀모형을 추정해 볼 수 있으나, 지질학적인 관점에서 공극을 또한 산사태 발생에 중요한 영향을 미치는 변수라고 사료되어 이를 포함하여 총 7개의 변수를 이용하여 로지스틱 회귀모형을 추정하였다.

추정된 회귀모형의 신뢰성 검증을 위해 정밀 조사 지역 중 경기북부의 장흥과 진접지역을 대상으로 산사태 발생가능성을 정량적으로 평가해 보았다. 산사태 발생가능성을 정량적으로 산출한 결과, 연구지역 내 발생한 산사태의 대부분이 90%이상의 산사태 발생 확률을 나타내었고, 기 발생 산사태 중 5개의 산사태만이 50%미만의 발생확률을 보였다. 또한, 산사태 미 발생지점에 대한 발생가능성을 도출한 결과, 4개 지점을 제외한 모든 지점이 15%미만의 산사태 발생확률을 나타내었다. 이 결과는 전술한 산사태 발생인자의 가중치 설정이 매우 정확하며, 높은 산사태 예측률을 가짐을 의미한다.

사 사

이 연구는 과학기술부에서 시행하는 특정연구개발사업의 하나인 '자연재해방지기술개발사업'으로 수행된 것이다.

참고문헌

김원영, 채병곤, 김경수, 기원서, 조용찬, 최영섭, 이사로, 이봉주, 2000, 산사태 예측 및 방지기술 연구, 과학기술부, KR-00-(T)-09.
 김원영, 채병곤, 김경수, 조용찬, 최영섭, 이춘오, 이철우, 김구영, 2003, 산사태 예측 및 방지기술 연구, 과학기술부, KR-03-(T)-03.
 채병곤, 김원영, 조용찬, 김경수, 이춘오, 최영섭, 이

문세, 2003, 지질별 산사태 정량적 예측모델, 2003 대한지질공학회 학술발표회, pp.11~19.
 성응현, 2001, 응용로지스틱 회귀분석, 탐진, 서울, 334p.
 Dikau R., Brunnsden, D., Schrott, L. and Ibsen, M.-L., 1996, Landslide Recognition, John Wiley and Sons, New York, pp.149~188.
 UNESCO Working Party on Landslide Inventory, 1990, A suggested method for reporting a landslide, Bull. of IAEG, no. 41, pp.5~12.
 EPOCH(European Community Programme), 1993, Temporal occurrence and forecasting of landslides in the European Community. 3 Volumes, Contract no. 90 0025, Brunnsden, 1985,
 Hutchinson, J. N., 1988, Morphological and geotechnical parameters of landslides in relation to geology and hydrology, In Landslides Proc. 5th Int. Symp. on Landslides, vol. 1, pp.3~35.
 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.
 千木良雅弘, 1997, 風化と崩壊, 近未来社, 名古屋, pp.49-87.

투 고 일 2004년 5월 18일

심 사 일 2004년 5월 19일

심사완료일 2004년 6월 15일

채병곤
 한국지질자원연구원 지질환경재해연구부
 305-350, 대전시 유성구 가정동 30
 Tel: 042-868-3052
 Fax: 042-861-9723
 E-mail: bgchae@kigam.re.kr

김원영

한국지질자원연구원 지질환경재해연구부
305-350, 대전시 유성구 가정동 30
Tel: 042-868-3345
Fax: 042-861-9723
E-mail: kwy@kigam.re.kr

조용찬

한국지질자원연구원 지질환경재해연구부
305-350, 대전시 유성구 가정동 30
Tel: 042-868-3059
Fax: 042-861-9723
E-mail: choych@kigam.re.kr

김경수

한국지질자원연구원 지질환경재해연구부
305-350, 대전시 유성구 가정동 30
Tel: 042-868-3054
Fax: 042-861-9723
E-mail: kks@kigam.re.kr

이춘오

한국지질자원연구원 지질환경재해연구부
305-350, 대전시 유성구 가정동 30
Tel: 042-868-3131
Fax: 042-861-9723
E-mail: lco@kigam.re.kr

최영섭

한국지질자원연구원 지질환경재해연구부
305-350, 대전시 유성구 가정동 30
Tel: 042-868-3117
Fax: 042-861-9723
E-mail: yschoi@kigam.re.kr