

## 정준상관 기반의 수량화분석에 의한 산사태 취약성 평가기법 제안

채병곤<sup>1</sup> · 서용석<sup>2\*</sup>

<sup>1</sup>한국지질자원연구원 지구환경연구본부, <sup>2</sup>충북대학교 지구환경과학과

### Suggestion of an Evaluation Chart for Landslide Susceptibility using a Quantification Analysis based on Canonical Correlation

Byung-Gon Chae<sup>1</sup> and Yong-Seok Seo<sup>2\*</sup>

<sup>1</sup>Geologic Environment Research Division, Korea Institute of Geoscience and Mineral Resources, Daejeon 305-350, Korea

<sup>2</sup>Dept. of Earth and Environmental Sciences, Chung Buk National University, Cheongju 361-763, Korea

Probabilistic prediction methods of landslides which have been developed in recent can be reliable with premise of detailed survey and analysis based on deep and special knowledge. However, landslide susceptibility should also be analyzed with some reliable and simple methods by various people such as government officials and engineering geologists who do not have deep statistical knowledge at the moment of hazards. Therefore, this study suggests an evaluation chart of landslide susceptibility with high reliability drawn by accurate statistical approaches, which the chart can be understood easily and utilized for both specialists and non-specialists. The evaluation chart was developed by a quantification method based on canonical correlation analysis using the data of geology, topography, and soil property of landslides in Korea. This study analyzed field data and laboratory test results and determined influential factors and rating values of each factor. The quantification analysis result shows that slope angle has the highest significance among the factors and elevation, permeability coefficient, porosity, lithology, and dry density are important in descending order. Based on the score assigned to each evaluation factor, an evaluation chart of landslide susceptibility was developed with rating values in each class of a factor. It is possible for an analyst to identify susceptibility degree of a landslide by checking each property of an evaluation factor and calculating sum of the rating values. This result can also be used to draw landslide susceptibility maps based on GIS techniques.

**Key words :** landslide prediction, canonical correlation analysis, quantification analysis, influential factor, evaluation chart of landslide susceptibility

최근 다양하게 제시되고 있는 확률론적 방법에 의한 산사태 예측기법의 경우 전문적 지식을 기반으로 조사 및 분석이 이루어질 경우에만 분석결과의 신뢰성을 확보할 수 있다. 그러나 재해 발생상황에서는 통계분석을 통한 산사태 예측의 전문가뿐만 아니라 공무원, 지질공학자 등 통계적 전문지식을 갖지 않은 재해분야 담당자도 신뢰성 있고 간편한 방법으로 산사태 취약성을 해석할 수 있어야 한다. 따라서 본 논문은 전문가는 물론 비전문가도 쉽게 의미를 이해하고 활용할 수 있으면서도 정확한 분석을 통한 통계적 접근으로 신뢰성 높은 산사태 취약성 평가표를 개발하여 제안하고자 하였다. 이를 위해 기존에 국내에서 산사태가 집중적으로 발생한 지역의 지질, 지형, 토질자료를 토대로 산사태 정준상관분석을 통한 수량화 기법을 이용하여 산사태 취약성 평가표를 개발하였다. 산사태의 현장자료와 실내시험자료를 바탕으로 통계분석을 실시하고, 그 결과를 토대로 영향인자 선정 및 인자별 급간 값을 설정한 것이다. 수량화 분석결과 산사태를 발생시키는 여러 인자 중 사면경사가 가장 큰 중요도를 가지며, 고도, 투수계수, 간극율, 암질, 건조밀도의 순서로 큰 영향을 미치는 것으로 나타났다. 각 평가항목별로 결정된 점수를 기준으로 평가항목 각각의 세부등급에 대한 점수를 할당하여 산사태재해 취약성 평가표를 개발하였다. 산사태재해 취약성 평가표를 이용하여 평가자는 평가대상 지점에 대해 각 평가항목별 해당 속성, 즉 세부등급을 선택하고, 선택된 각 속성별 평가점수를 더하면

\*Corresponding author: ysseo@chungbuk.ac.kr

산사태 취약성을 점수로 신속하게 파악할 수 있다. 또한, 이 결과를 토대로 GIS 기법을 이용한 산사태 예측지도 또는 취약성지도 등을 작성하여 활용할 수 있다.

**주요어** : 산사태 예측, 정준상관분석, 수량화 분석, 영향인자, 취약성 평가표

## 1. 서 론

2000년대 이후부터 국내 뿐만 아니라 해외에서는 산사태 발생가능성 예측과 관련하여 다양한 연구를 수행하면서 과거의 결정론적 방법에서 탈피하여 확률론적 방법을 이용한 예측기법을 집중적으로 연구해 오고 있다 (Chae *et al.*, 2004a; Chae *et al.*, 2004b; Dai *et al.*, 2001; Davis *et al.*, 2006; Lee *et al.*, 2004; Nandi and Shakoor, 2009; Van Den Eeckhaut *et al.*, 2006; Yilmaz, 2009). 이러한 연구는 대부분 이변량 분석기법 또는 다변량 분석기법을 사용하여 산사태 발생에 영향을 미치는 다양한 인자를 선별하고, 이를 토대로 산사태 취약성 또는 발생가능성을 예측하는 식을 도출하였다. 이와 같은 방법은 여러 변수를 동시에 고려하고, 각 변수들 간의 영향력을 서로 평가할 수 있어 결정론적 방법에서 불가능했던 다양한 조건 또는 경우를 고려할 수 있게 되었다. 이외에도 인공지능망기법을 동원한 산사태 예측방법도 개발되어 다양한 지역에서 그 성능을 검증하기도 하였다 (Gomez and Kavzoglu, 2005; Lee *et al.*, 2003; Kavzoglu, 2001). 그러나 이러한 예측방법은 대부분 전문적인 지식을 기반으로 자료를 분석하고 그 결과에 대한 해석이 이루어질 경우에만 신뢰성이 보장된다. 따라서 해당분야의 전문가가 아닌 경우에는 정확한 산사태 예측을 하기 어려운 문제점이 발생한다.

최근의 기상이변에 따른 급작스런 국지성 집중 호우와 이로 인한 급격한 산사태의 발생은 과거에 비해 상당히 짧은 시간동안 이루어지기 때문에 복잡한 절차와 전문가적 접근방법에 의한 예측기법에만 의존할 경우 제해 상황에 효과적으로 대응하기 어렵다. 따라서 재해담당 공무원 또는 기관이 비교적 쉽게 활용할 수 있는 산사태 예측 및 평가방법이 필요하다. 이러한 측면에서 기존에 산림청, 소방방재청 등에서 제시한 몇 가지의 산사태 예측방법이 있다. 산림청의 “산사태 위험지 판정표”는 사면의 기하형태, 모암, 토심, 임상 등의 인자를 기준으로 각 인자의 급간에 따라 값을 설정하고, 이들 값의 합에 따라 산사태의 발생가능성을 평가하는 것이다. 기본적으로 산정된 점수를 바탕으로 5개 항목에 대한 점수보정을 실시하여 최종적인 산사태 발

생가능성을 산정하게 된다. 이 판정표는 지질학 또는 지반공학적인 전문지식이 없더라도 산사태의 가능성을 비교적 쉽게 평가할 수 있는 장점이 있다. 따라서 넓은 지역을 대상으로 짧은 시간에 산사태 발생가능성을 평가할 수 있다. 그러나 이 판정표는 기본적으로 산사태 물질인 토층의 지질학적·토질역학적 특성이 고려되어 있지 않다. 또한, 점수보정 항목 중 일부는 조사자의 주관이 크게 개입될 가능성이 있는 항목이 있으며, 이럴 경우 평가결과에 오류가 발생할 수도 있을 것으로 생각된다. 소방방재청에서 제안한 “급경사지 위험 평가표”의 경우는 각 항목의 급간값이 일정한 분포 범위로만 할당되어 있어 최종적인 평가결과에 신뢰성에 문제가 있다. 뿐만 아니라 연간 강우량, 강우강도, 강우 지속일 등 강우에 대한 영향을 지나치게 많이 강조하고 있어 다른 항목과의 중요도 또는 가중치에 문제가 있음이 사실이다.

이와 같은 상황에서 전문가는 물론 비전문가도 쉽게 의미를 이해하고 활용할 수 있으면서도 정확한 분석을 통한 통계적 접근으로 신뢰성 높은 산사태 예측 평가표가 제시되어야 한다. 따라서 본 연구에서는 기존 급경사지재해 평가기법을 보완하고, 정확한 평가결과를 도출하기 위하여 응용통계분석을 기반으로 새로운 산사태 취약성 평가표를 개발하고자 하였다. 이 평가표는 정준상관분석을 통한 수량화 기법을 이용하여 개발한 것으로서, 그동안 국내에서 발생한 산사태의 현장 자료와 실내시험자료를 바탕으로 통계분석을 실시하고, 그 결과를 토대로 영향인자 선정 및 인자별 급간 값을 설정한 것이다. 이 평가표를 이용할 경우 분석대상 지역의 산사태 발생가능성을 점수 형태로 도출할 수 있으며, 이를 토대로 GIS 기법을 이용한 산사태 예측지도 또는 취약성지도 등을 작성하여 활용할 수 있다.

## 2. 정준상관분석과 수량화

정준상관분석은 다중회귀분석과 상관분석의 확장으로 생각하면 쉽게 이해할 수 있다. 다중선형회귀분석에서는 모형에 포함된 여러 독립변수들의 최적선형결합을 이용하여 하나의 종속변수의 변량을 설명한다. 추정된 다중회귀함수의 정도를 측정하는 방법으로 흔히

사용되는 다중결정계수는 종속변수의 관측치와 모형에 의하여 추정된 추정치 사이의 단순상관계수이다. 따라서 다중회귀분석에서의 관심사는 여러 개의 독립변수들과 하나의 종속변수 사이의 선형적인 함수관계를 살펴보는 데 있다. 반면 정준상관분석에서는 여러 변수들을 두 변수집단인 종속변수집단(기준변수집단)과 독립변수집단(설명변수집단)으로 분류한 후, 두 변수집단 사이의 상관구조를 가장 잘 설명할 수 있는 변수들의 선형결합을 찾는 것이다. 특히 정준상관분석은 종속변수집단의 변수들이 서로 상관관계에 있을 때 매우 유용하게 사용될 수 있다.

2.1. 정준변수와 정준상관

본 절에서는 모집단 정준상관모형에서 정준변수와 정준상관계수가 어떻게 유도되는지를 설명한다.  $p+q$ 개 확률변수들을 두 개 집단으로 분류할 때  $p$ 개 확률변수로 구성된 종속변수 집단을 확률벡터  $Y$ 로 표시하고,  $q$ 개 확률변수로 구성된 독립변수집단을  $X$ 로 표시하기로 한다. 이론적 전개를 위하여  $p \leq q$ 라고 가정한다. 확률벡터  $Y$ 와  $X$ 의 평균벡터와 분산공분산행렬은 각각 다음과 같이 표시된다.

$$\begin{aligned}
 Y &= (Y_1, Y_2, \dots, Y_p)', X = (X_1, X_2, \dots, X_q)' \\
 E(Y) &= \mu_y = (\mu_{y1}, \mu_{y2}, \dots, \mu_{yp})' \\
 E(X) &= \mu_x = (\mu_{x1}, \mu_{x2}, \dots, \mu_{xp})' \\
 Cov(Y) &= \sum_{yy} = E\{(Y - \mu_y)(Y - \mu_y)'\} \\
 Cov(X) &= \sum_{xx} = E\{(X - \mu_x)(X - \mu_x)'\} \\
 Cov(Y, X) &= \sum_{yx} = \sum'_{xy} = E\{(Y - \mu_y)(X - \mu_x)'\} \quad (1)
 \end{aligned}$$

위 식에서  $\sum_{yy}$ 와  $\sum_{xx}$ 는 각 변수집단 내 분산공분산행렬을 의미하고,  $\sum_{yx}$ 는 두 변수집단 간 공분산행렬을 의미한다. 두 집단을 결합한  $p+q$ 차원 확률벡터  $U = (Y, X)$ 를 고려하면 확률벡터  $U$ 의 평균벡터  $\mu_u$ 와 분산공분산행렬  $\sum_{uu}$ 은 다음과 같이 분할되어 표시된다.

$$\begin{aligned}
 \mu_u &= E(U) = \begin{bmatrix} E(Y) \\ E(X) \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix} \\
 \sum_{uu} &= E\{(U - \mu_u)(U - \mu_u)'\} \\
 &= \begin{bmatrix} Cov(Y) & \vdots & Cov(Y, X) \\ \vdots & \ddots & \vdots \\ Cov(X, Y) & \vdots & Cov(X) \end{bmatrix} \\
 &= \begin{bmatrix} \sum_{yy} & \vdots & \sum_{yx} \\ \vdots & \ddots & \vdots \\ \sum_{xy} & \vdots & \sum_{xx} \end{bmatrix} \quad (2)
 \end{aligned}$$

두 변수 집단 사이의 연관관계는 두 집단 간의 공분산행렬  $\sum_{yx}$ 의  $(p \times q)$ 개 원소를 살펴보면 대략적으로 파악할 수 있을 것이다. 그러나 각 집단 변수의 수가 크다면  $\sum_{yx}$ 의 원소들을 해석한다는 것은 거의 불가능하다. 이러한 경우 정준상관분석에서는  $\sum_{yx}$ 의  $(p \times q)$ 개 공분산들보다는 소수개 공분산(혹은 상관)들을 이용하여  $Y$ 와  $X$ 의 연관성을 요약하게 된다.

$Y$ 와  $X$ 의 연관관계는 각 변수집단에서 선택되는 변수들의 선형결합으로 나타나게 된다. 각 변수집단에서  $p$ 차원 계수벡터  $a$ 와  $q$ 차원 계수벡터  $b$ 를 이용한 선형결합  $W$ 와  $V$ 를 다음 식 (3)과 같이 생각해 보자.

$$\begin{aligned}
 W &= aY = a_1 Y_1 + a_2 Y_2 + \dots + a_p Y_p \\
 V &= bX = b_1 X_1 + b_2 X_2 + \dots + b_q X_q \quad (3)
 \end{aligned}$$

여기서 선형결합에 사용된 계수  $a, b$ 는 선형결합  $W$ 와  $V$ 에서 사용된 각 변수들의 가중치를 의미한다. 선형결합  $W, V$ 의 분산과 공분산을 구하면

$$\begin{aligned}
 Var(W) &= a' Cov(Y) a = a' \sum_{yy} a \\
 Var(V) &= b' Cov(X) b = b' \sum_{xx} b \\
 Cov(W, V) &= a' Cov(Y, X) b = a' \sum_{yx} b \quad (4)
 \end{aligned}$$

이 된다. 또한 확률변수  $W$ 와  $V$ 의 단순상관계수  $\rho(W, V)$ 는

$$\begin{aligned}
 \rho(W, V) &= \frac{Cov(W, V)}{\{Var(W)Var(V)\}^{1/2}} \\
 &= \frac{a' \sum_{yx} b}{\{(a' \sum_{yy} a)(b' \sum_{xx} b)\}^{1/2}} \quad (5)
 \end{aligned}$$

이 된다. 정준상관분석에서는  $Y$ 와  $X$ 에 대한 많은 선형결합  $W$ 와  $V$ 중에서 식 (5)의 단순상관계수를 최대화시킬 수 있는 계수벡터  $a$ 와  $b$ 를 찾는 것이다. 상관계수  $\rho(W, V)$ 는 계수벡터  $a$ 와  $b$ 의 함수이며, 척도의 불변성이 있기 때문에  $W$ 와  $V$ 의 분포를 평균이 0이고 분산이 1이 되도록 표준화 시켜도 일반성을 잃지 않게 될 것이다.

$$\begin{aligned}
 E(W) &= E(V) = 0 \\
 Var(W) &= Var(V) = a' \sum_{yy} a = b' \sum_{xx} b = 1 \quad (6)
 \end{aligned}$$

식 (5)의 단순상관계수를 최대화시키는 계수벡터  $a, b$ 를 정준계수벡터라고 하고, 구해진 정준계수를 선형결합계수로 상용한  $W$ 와  $V$ 를 정준변수라고 한다.

## 2.2. 정준구조행렬

정준변수는 변수들의 선형결합으로 이루어진 인공변수이기 때문에 특정한 의미를 부여하기가 쉽지 않다. 정준변수에서 정준계수의 해석은 정준계수의 크기와 부호에 따라 가중치의 의미를 부여하게 되며, 가중치의 크기는 각 변수가 정준변수에 미치는 공헌도를 의미하게 된다. 그러나 각 변수 집단 내에서 변수들 사이에 높은 다중공선성이 존재한다면 정준계수의 크기와 부호의 의미를 신뢰성 있게 해석할 수 없게 된다. 따라서 정준계수와 더불어 다음에 설명할 정준적재와 정준교차적재를 계산하여 해석하여야 한다.

정준적재(canonical loading)는 각 변수집단의 정준변수와 해당 변수집단 내 변수 사이의 단순상관관계를 의미한다. 만약 정준적재가 크다면 그 변수가 정준변수에 미치는 영향이 크다고 볼 수 있다. 또한 어떤 변수집단의 변수들이 다른 변수집단의 정준변수와 어떤 관계가 있는지를 살펴보기 위해서는 정준교차적재(canonical cross-loadings)를 계산하여야 한다. 만약 정준교차적재가 크다면 어떤 변수집단의 변수가 다른 변수집단의 정준변수에 미치는 영향이 크다고 볼 수 있다.

정준계수와 정준적재는 어느 한 집단에서 유도된 정준변수와 변수들 사이의 연관성을 평가하는데 사용될 수 있다. 다시 설명하면 정준계수는 다중회귀분석에서 회귀계수와 유사한 의미를 갖는다. 왜냐하면 각 변수들이 집단 내 정준변수들의 변동을 설명할 수 있는 공헌도를 의미하기 때문이다. 그러나 정준계수들은 집단 내 변수들 사이에 다중공선성이 존재할 때에는 매우 불안정하게 된다. 예를 들면 어느 변수가 다른 여러 변수와 다중공선성이 존재한다면, 그 변수는 다른 여러 변수들과 중복된 변수를 갖고 있기 때문에 정준계수는 매우 작은 값을 갖거나 혹은 기대되는 부호와 다른 부호를 갖게 되는 경우가 종종 발생하게 된다. 다중공선성 때문에 발생하는 정준계수의 불안정성 이외에도 정준계수를 해석할 때 특별한 의미를 부여하기가 쉽지 않다. 왜냐하면 다중회귀분석에서는 변수들의 관측이 가능하지만, 정준변수는 집단 내 변수들의 선형결합으로 이루어진 인공변수이기 때문이다.

따라서 정준상관구조를 분석하기 위해서는 정준계수보다 정준적재를 사용하는 것이 바람직하다. 왜냐하면 정준적재는 어느 한 집단 내 정준변수들과 변수들 사이의 상관계수이기 때문에 정준변수에 사용된 개별 변수들의 공헌도를 설명할 수 있기 때문이다. 그러나 정준적재는 정준계수에 비해 다중공선성의 영향을 작게

받지만 다음과 같은 약점이 있다. 첫째, 정준적재는 잠재적으로 정준계수와 마찬가지로 불안정한 결과를 보인다. 둘째, 정준적재가 안정적이라 하더라도 한 집단 내 정준변수와 변수들 사이의 상대적인 상관정도를 비교할 수 있을 뿐이지 한 집단 내 정준변수와 다른 집단의 변수들 사이의 상관구조를 알 수 없다는 약점을 가지고 있다. 따라서 이와 같은 단점을 극복하기 위해 정준교차적재를 함께 사용하여 정준상관구조를 해석하는 것이 바람직하다고 판단된다.

## 3. 정준상관분석 기반의 수량화를 통한 산사태 예측

위에서 소개한 정준상관분석 과정을 산사태 자료에 대해 적용한다. 분석에 사용될 자료는 4개 지역(포천~서울, 상주, 포항, 속초)별 자료와 이들 자료를 지질별로 통합한 자료에 대해 적용하였다. 각 케이스별 산사태 측정자료에 대해 수량화 방법을 적용하는 절차를 설명하면 다음과 같다.

**[단계1]** 산사태 발생여부를 반응변수( $Y$ )로 하고, 이에 영향을 미치는 25개의 변수 즉, Lithology( $X_1$ ), Weathering( $X_2$ ), Elevation( $X_3$ ), ..., SiltClay( $X_{25}$ )를 설명변수로 설정한다. 이들 중 Lithology( $X_1$ ), Weathering( $X_2$ ), USCS( $X_{18}$ )은 범주형 변수이며, 나머지 23개는 연속형 변수이다.

**[단계2]** 정준상관분석을 위해 모든 범주형 설명변수들을 가변수를 이용하여 나타내고, 연속형변수와 결합하여 모든 설명변수들의 가변수 그룹을  $X=(X_1, \dots, X_q)$ 으로 정의한다. 반응변수는 범주가 2개 이므로(발생=1, 미발생=0) 가변수 그룹을  $Y$ 로 둔다.

**[단계3]** 정준상관분석을 실시하여 정준변수와 정준계수를 구한다. 즉, 다음의 선형결합식

$$W=aY \quad (7)$$

$$V=bX=b_1X_1+b_2X_2+\dots+b_qX_q \quad (8)$$

에서, 제1 정준변수쌍이라 불리는  $(\hat{W}_1, \hat{V}_1)$ 은  $W$ 와  $V$  간의 단순상관계수를 최대화 시키는 변수집단내의 선형결합으로 각각의 분산이 단위분산이 되도록 구성한다.

**[단계4]** 각 범주형 설명변수에 대해 각 범주별 회귀계수의 추정치를 구하고, 이 값에 표본정준상관계수를 곱하여 1차 수량화값을 계산한다. 연속형 변수에 대해서는 회귀계수의 값을 그대로 1차 수량화값으로 한다.

[단계5] 1차 수량화값을 이용하여 편상관을 구한다. 여기서 편상관은 1차 수량화 결과로부터 모든 질적변량을 양적변량으로 대체한 후 다중회귀를 통해 구해진다. 편상관을 사용하는 수량화 과정은 실제 자료분석 과정에서 설명하기로 한다.

[단계6] 계산된 각 변수별 편상관값을 상대적 중요도 즉, 퍼센트(%)로 나타내어 최종의 수량화값(가중치)을 구한다.

### 3.1. 분석자료 및 변수 설명

Table 1은 수량화 분석에 사용된 자료를 요약한 것으로서, 자료의 수를 충분히 확보하기 위해 결측치가 많은 변수는 사전에 제거하였다. 지질 및 지역별로 분석에 사용된 자료를 정리하면 Table 1과 같다.

Table 2는 분석에 사용된 변수들을 요약한 것이다. 이 표에서 각 지역 및 지질별 자료분석 과정에서 분석에 포함된 변수와 제외된 변수를 나타낸 것이며, 분석

**Table 1.** Basic information of data number for quantification analysis

Item	Total number of raw data	After elimination of data absent variables			
		No. of raw data	Data No. on slided area	Data No. of not-slided area	
Geology	Lithology	1214	310	121	189
Area	Gyeonggi	97	53	26	27
	Pohang	91	81	39	42
	Sangju	95	55	29	26
	Sokcho	135	31	10	20

**Table 2.** Variables for the quantification analysis

Category	Variable	Variables used for each analysis				
		Geology	Area 1 (Gyeonggi)	Area 2 (Pohang)	Area 3 (Sangju)	Area 4 (Sokcho)
Target variable	Slided or Not-slided	○	○	○	○	○
Input variable	Lithology	○	○	○	○	○
	Weathering	×	×	×	×	×
	Elevation	○	○	○	○	○
	Slope Direction	×	○	×	×	×
	Slope Angle	○	○	○	○	×
	Triggering Position	×	×	×	×	×
	Length	×	×	×	×	×
	Width	×	×	×	×	×
	Thickness	×	×	×	×	×
	Specific Gravity	○	○	○	×	○
	Moisture Content	○	○	○	×	○
	Void ratio	○	○	○	×	○
	Porosity	○	○	○	○	○
	Saturation	○	○	○	×	○
	Wet Density	○	○	○	×	○
	Sat_Density	○	○	○	×	○
	Dry Density	○	○	○	○	○
USCS (*)	○	○	○	○	○	
Permeability	○	○	○	○	○	
Liquid Limit	×	×	×	×	×	
Plastic Limit	×	×	×	×	×	
Plast_Index	×	×	×	×	×	
Gravel	×	×	×	×	○	
Sand	×	×	×	×	○	
Silt Clay	×	×	×	×	○	

에 제외된 변수 가운데는 중요한 변수가 포함될 수 있음에 유의할 필요가 있다. 예를 들어, 속초지역의 경우 사면경사(Slope Angle) 변수가 제외되었음을 알 수 있다. 사면경사 변수는 산사태자료에서 일반적으로 중요한 변수로 취급되어지므로, 결과의 해석 시에 타 지역의 결과를 고려하여 해석할 필요가 있다.

### 3.2. 정준상관 및 수량화 분석 결과

지질별 자료(n=310)에 대해 정준상관분석을 수행한 결과는 다음과 같다. Table 3에서 정준계수는 정준상관분석을 통해 구해지며, 편상관은 구해진 정준계수를 이용하여 모든 질적변량을 양적변량으로 수량화 한 뒤, 양적 설명변수와 양적 반응변수 간의 다중회귀분석을

**Table 3.** Results of canonical correlation analysis for geologic data

Variable	Canonical coefficient	Partial correlation	
Landslide	1(slided)	2.000	
	0(not-slided)	0.000	
Lithology	Gneiss=1	-0.175	0.148
	Mudstone=2	-0.416	0.148
	Schist=3	0.173	0.148
	Granite=4	0.059	0.148
	Tuff=5	0.000	0.148
Elevation	-0.002	0.219	
Slope Angle	0.075	0.625	
Gravity	2.966	0.126	
Content	0.060	0.084	
Void ratio	1.594	0.145	
Porosity	-0.023	0.184	
Saturation	-0.046	0.138	
Wet Density	1.807	0.032	
Sat_Density	1.376	0.000	
Permeability	-0.464	0.217	
Dry density	-0.038	0.000	
USCS	sm-sc=1	0.782	0.308
	sc=2	0.296	0.308
	sm=3	-0.940	0.308
	sw-sm=4	1.062	0.308
	cl=5	-0.521	0.308
	sw=6	0.555	0.308
	sw-rc=7	.	0.308
	sw-sc=8	0.350	0.308
	sp=9	0.808	0.308
	sp-sc=10	0.000	0.308
Canonical correlation	0.808		

통해 구해진다. Table 3에서 제공되는 편상관의 크기를 상대적인 크기로 나타내어, 반응변수에 미치는 중요도를 계산한 결과는 Table 4와 같다.

지질별 자료에 대해 수량화 분석을 실시하였다. 동일한 방법을 지역별 자료에 적용하고, 이를 요약하면 Table 5를 얻게 된다. Table 5에서는 지질 및 지역별로 산사태 발생에 미치는 중요 인자를 순위로 표현한 것이다. 이 결과들을 종합적으로 고려하여 선정된 중요 인자에 대해 추후 산사태 취약성 평가표 작성의 기초자료로 활용하게 된다.

위의 결과를 토대로 산사태 발생에 영향을 미치는 중요 인자를 6개의 인자로 선정하였다(Table 6). 이 과정에서는 수량화 분석에서의 결과를 토대로 관련 전문가의 의견을 고려하여 최종 결정되었다. 중요도 점수의 산정은 지질별 분석 결과에서 선정된 6개의 인자만을 고려하여 100점 만점으로 환산한 것이다.

### 4. 산사태 취약성 평가표 개발

이 절에서는 앞서 선택된 6개의 주요변수에 대한 기초 통계치를 제시하고, 각 변수별로 반응변수(산사태 발생여부)와의 관계를 모형화 하였다. 모형화 과정에는 직선회귀(linear regression)와 로지스틱회귀(logistic regression)가 사용되었다. 예를 들어 암종(lithology)의 경우 Table 7은 국내에서 산사태가 집중적으로 발생한 주요 암질의 유형(화강암, 편마암, 제3기 이암)에 따른 산사태 발생 확률을 나타낸다.

**Table 4.** Calculation of importance of variables for geologic data

Variable	Partial correlation	Importance	Rank
Lithology	0.148	6.6%	6
Elevation	0.219	9.8%	3
Slope Angle	0.625	28.1%	1
Specific Gravity	0.126	5.7%	9
Moisture Content	0.084	3.8%	10
Void ratio	0.145	6.5%	7
Porosity	0.184	8.3%	5
Saturation	0.138	6.2%	8
Wet Density	0.032	1.4%	11
Sat_Density	0.000	0.0%	12
Dry Density	0.000	0.0%	12
USCS	0.308	13.8%	2
Permeability	0.217	9.7%	4
Total	2.226	100.0%	

**Table 5.** Importance ranks of each variable for geology and each area

Variable	Geology	Area 1(Gyeonggi)	Area 2(Pohang)	Area 3(Sangju)	Area 4(Sokcho)
Lithology	6	-	13	2	9
Elevation	3	7	3	6	3
Slope Angle	1	2	1	1	-
Specific Gravity	9	12	6	-	4
Moisture Content	10	8	11	-	8
Void ratio	7	11	12	-	5
Porosity	5	8	9	5	9
Saturation	8	8	8	-	6
Wet Density	11	6	7	-	7
Sat_Density	12	4	5	-	9
Dry Density	12	5	10	3	9
USCS	2	1	4	4	1
Permeability	4	3	2	7	2
Gravel	-	-	-	-	9
Sand	-	-	-	-	9
Silt Clay	-	-	-	-	9
Slope Direction	-	13	-	-	-
Reliability	High	High	High	High	High

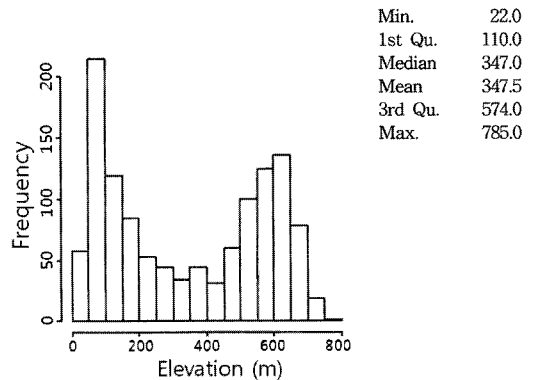
**Table 6.** Result of influential variable selection by quantification analysis

Variable	Score(Importance)	Rank
A. Lithology	10(9.7)%	5
B. Elevation(m)	15(14.4)%	2
C. Slope angle(°)	40(41.2)%	1
D. Porosity(%)	12(12.2)%	4
E. Dry density(t/m <sup>3</sup> )	8(8.4)%	6
F. Permeability(cm/s)	15(14.2)%	3
Total	100%	

**Table 7.** Landslide probability dependent on lithology

Lithology(n=1057)	Probability : P(y=1)
Granite(n=833)	0.911
Gneiss(n=139)	0.806
Mudstone(n=85)	0.529

한편, 고도(elevation)에 대한 분석결과의 히스토그램과 요약된 통계치는 Fig. 1과 같다. Fig. 2는 자료에 대한 여러 가지 정보와 모형화 분석결과를 한꺼번에 나타낸 그림이다. 먼저, y축의 0과 1 상에 점들은 고도 변수에 따른 반응변수값(1또는 0)을 나타내며, 고도 변수의 구간을 15개로 나눈 뒤 각 구간별 발생확률을 15개의 점으로 나타내었으며, 원자료에 대해 직선회귀와 로지스틱회귀를 적합한 곡선을 그린 것이다. 우측에 제시된 표는 원 자료를 15개의 등급으로 나눈 결과



**Fig. 1.** Histogram and summarized statistics for elevation.

를 나타낸다.

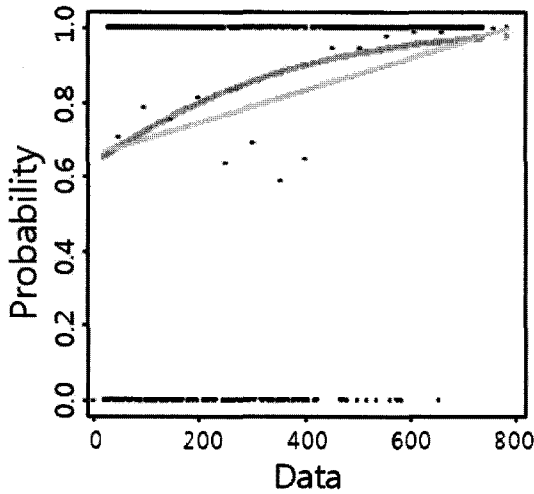
원 자료에 대해 로지스틱과 직선회귀를 적합한 결과(KIGAM, 2008)로부터 산사태 발생확률에 대한 다음의 추정식을 얻을 수 있다.

로지스틱회귀:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right)=0.5344+0.0041 \cdot Elevation$$

또는 
$$\hat{p}=\frac{\exp(0.5344+0.0041 \cdot Elevation)}{1+\exp(0.5344+0.0041 \cdot Elevation)} \quad (9)$$

직선회귀: 
$$\hat{p}=0.6590+0.0004 \cdot Elevation \quad (10)$$



LCL	MID	UCL	N	PROB
22.0	47.4	72.8	164	0.707
72.8	98.3	123.7	174	0.787
123.7	149.1	174.6	112	0.758
174.6	200.0	225.4	60	0.816
225.4	250.9	276.3	42	0.642
276.3	301.7	327.2	39	0.692
327.2	352.6	378.0	44	0.590
378.0	403.5	428.9	37	0.648
428.9	454.3	479.8	38	0.947
479.8	505.2	530.6	80	0.950
530.6	556.1	581.5	136	0.977
581.5	606.9	632.4	128	0.992
632.4	657.8	683.2	109	0.990
683.2	708.7	734.1	34	1.000
734.1	759.5	785.0	6	1.000

Fig. 2. Landslide probability of each class of elevation and goodness of fit model.

Table 8. Calculation of percentile for elevation

Percentile	Logistic regression		Linear regression	
	Inferred probability	x	Inferred probability	x
100	0.9764	777.6201	1.0000	852.4950
95	0.9624	660.5310	0.9486	724.1171
90	0.9583	633.9837	0.9370	695.0094
85	0.9549	614.1718	0.9283	673.2872
80	0.9513	594.3599	0.9196	651.5651
75	0.9461	568.6026	0.9083	623.3263
70	0.9409	544.8301	0.8979	597.2597
65	0.9351	520.3621	0.8872	570.4328
60	0.9244	480.4414	0.8697	526.6627
55	0.9020	411.0997	0.8393	450.6352
50	0.8748	343.7392	0.8097	376.7799
45	0.8312	258.4488	0.7723	283.2660
40	0.7968	202.8764	0.7479	222.3354
35	0.7678	161.3801	0.7297	176.8383
30	0.7462	132.7028	0.7172	145.3955
25	0.7273	108.9681	0.7067	119.3724
20	0.7109	89.1562	0.6981	97.6502
15	0.6983	74.2973	0.6915	81.3586
10	0.6896	64.3913	0.6872	70.4975
5	0.6782	51.5136	0.6816	56.3781
0	0.6511	21.7957	0.6685	23.7949

과는 Table 8과 같다. 즉, 직선회귀와 로지스틱회귀의 적합결과로부터 발생확률이 높은 상위 5%, 10%, 15%에 대응하는 변수값(분위수)을 구한 것이다. Table 8에서 첫 열의 백분위는 변수의 최소값, 최대값에서의 추정확률의 차이를 20등분한 것이다.

이러한 과정을 거쳐 산사태 취약성 평가표를 구성하였다. Table 9는 앞에서 구한 분위수 정보를 이용하여 각 변수별로 산사태 발생에 미치는 영향을 3~5등급으로 분류한 결과이다. 적합된 두 가지 모형(로지스틱과 직선 모형) 중 보다 나은 적합도를 보이는 모형을 사용하였다. 각 변수별로 적합된 모형으로부터의 확률변화가 뚜렷한 지점을 반영하여 등급을 나누었다.

### 5. 수량화분석 결과의 성능평가

평가표의 성능을 비교하기 위하여 본 연구에서 제안한 수량화분석의 평가표를 통해 각각의 자료별 산사태 취약성 점수를 산출하였다. 예를 들어, 어느 한 지역에 대한 관측 자료가 Table 10과 같이 나타난다고 할 때 산사태 취약성에 대한 점수를 구해보면 다음과 같아진다. 위의 표에서 주어졌던 예의 관측값들을 Table 9의 평가표에 적용시켜 점수를 계산하면 다음과 같이 주어진다.

또한, 위와 같은 방법으로 사면경사(n=1113), 간극율(n=313), 건조밀도(n=313), 투수계수(n=230)에 대해 각각 분석을 수행하였다(KIGAM, 2008).

앞서 적합된 두 가지 모형으로부터 각 변수별 분위수를 5% 단위로 구하였으며, 그 예로 고도에 대한 결

$$\begin{aligned}
 & (Lithology=Gneiss)+(Elevation=234) \\
 & +(Slope\ Angle=28)+(Porosity=18.4) \\
 & +(Dry\ density=1.58) \\
 & +(Permeability=2.44 \cdot 10^{-5}) \\
 & =9+5+40+0+0+15=69
 \end{aligned}
 \tag{11}$$



**Table 9.** Evaluation chart of landslide susceptibility by quantification analysis

Observation Item		Class and score				
A. Lithology	Class	Granite		Gneiss	Mudstone	
	Class No.	1		2	3	
	Score	10		9	6	
B. Elevation (m)	Class	431~785		20~225	226~430	
	Class No.	1		2	3	
	Score	15		10	5	
C. Slope angle (°)	Class	>27.1	23.1~27.0	19.1~23.0	15.1~19.0	<15.0
	Class No.	1	2	3	4	5
	Score	40	30	20	10	0
D. Porosity (%)	Class	>80.1	75.1~80.0	67.1~75.0	46.1~67.0	<46.0
	Class No.	1	2	3	4	5
	Score	12	9	6	3	0
E. Dry density (g/cm <sup>3</sup> )	Class	0.80~1.18	1.19~1.31	1.32~1.38	1.39~1.47	>1.48
	Class No.	1	2	3	4	5
	Score	8	6	4	2	0
F. Permeability (cm/s)	Class	<1*10 <sup>-3</sup>	1.1*10 <sup>-3</sup> ~2.5*10 <sup>-3</sup>	2.6*10 <sup>-3</sup> ~4.5*10 <sup>-3</sup>	4.5*10 <sup>-3</sup> ~6.0*10 <sup>-3</sup>	6.0*10 <sup>-3</sup> ~1.0*10 <sup>-2</sup>
	Class No.	1	2	3	4	5
	Score	15	11	7	3	0

**Table 10.** Observation values to calculate landslide susceptibility

Variable	Lithology	Elevation (m)	Slope Angle (°)	Porosity (%)	Dry density (g/cm <sup>3</sup> )	Water contents (%)	Permeability (cm/s)
Value	Gneiss	234	28	18.4	1.58	8.05	2.44×10 <sup>-5</sup>

**Table 11.** Categorized chart of quantified score

Quantified score	Not-slided	Slided	Probability
0-10	-	-	-
11-20	-	-	-
21-30	10	-	-
31-40	19	-	-
41-50	8	3	27.3
51-60	9	-	-
61-70	4	7	63.6
71-80	4	19	82.6
81-90	-	23	100.0
91-100	-	6	100.0
Total	54	58	51.8

10점 단위로 범주화 하면 Table 11과 같다. 이 결과를 보면 수량화점수가 높을수록 산사태가 더 많이 발생하는 것을 알 수 있다.

앞에서 이루어진 산사태 발생 취약성 평가 점수에 따른 산사태 발생의 예측확률에 대한 모형을 개발하였다. 수량화 점수에 대한 자료(n=112)로부터 구간별 변수(1:발생, 0:미발생)를 반응변수로 하는 로지스틱회귀 모형의 적합결과는 다음의 Table 12-14와 같다.

위의 Hosmer-Lemeshow 검정 결과 적합된 로지스틱회귀모형이 잘 적합이 되었으며(p<0.001), 적합결과 는 다음과 같다.

위의 식과 같은 방법으로 각 관측치에 대해 수량화 점수를 산출한다. 각 자료별로 계산된 수량화 점수를

$$\hat{\pi} = \frac{\exp(-9.0937+0.1460 \cdot \text{quantifiedscore})}{1+\exp(-9.0937+0.1460 \cdot \text{quantifiedscore})} \quad (12)$$

**Table 12.** Result of goodness of fit for logistic regression model (quantification analysis)

Criterion	Intercept only	Intercept and Covariates	Chi-Square for GF Test
-2 Log L	155.122	61.007	5.4027 with 8 DF (p < 0.7138)

**Table 13.** Inference result of maximum likelihood (quantification analysis)

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr>Chisq
Intercept	1	-9.0937	1.7321	27.5638	<.0001
Quantified score	1	0.1460	0.0263	30.7138	<.0001

**Table 14.** Inference result of Odds ratio (quantification analysis)

Effect	Point Estimate	95% Wald Confidence Limits	
Quantified score	1.157	1.099	1.219

수량화 점수에 대한 추정치가 0보다 크기 때문에 수량화 점수가 커질수록 산사태 취약성에 대한 예측확률  $\hat{\pi}$ 가 커진다. 산사태 자료에서 점수가 1점 늘어날 때마다 산사태가 발생할 오르는 1.157배( $=e^{0.1460}$ ), 약 15.7%씩 증가한다.

## 6. 결 론

본 연구에서는 기존 급경사지재해 평가기법을 보완하고, 정확한 평가결과를 도출하기 위하여 응용통계분석을 기반으로 새로운 산사태 취약성 평가표를 개발하고자 하였다. 이 평가표는 과거 수년간 전국을 대상으로 실시한 현장조사의 데이터베이스를 이용하여 정준상관분석을 이용한 수량화기법을 적용하여 작성되었다. 본 연구에서는 산사태조사 DB에 포함되어 있는 전국 1,214개 산사태 발생/미발생지역의 조사자료 중 데이터 수가 통계적으로 유의한 편마암, 화강암, 제3기 이암을 대상으로 하였다. 분석에 사용된 산사태인자는 암종, 고도, 사면경사, 비중, 습윤도, 간극비, 간극율, 포화도, 습윤밀도, 건조밀도, USCS, 투수계수 등 총 13개 이다. 이들이 산사태발생에 미치는 영향도를 분석한 결과를 바탕으로 영향도 순위가 높고, 토질특성을 대표할 수 있는 6개의 영향인자를 최종적으로 선정하여 평가표의 항목으로 설정하였다.

이들 6개 선정요소에 대하여 각 요소별로 산사태발생에 미치는 영향도를 로지스틱 모형을 통하여 각 인자의 세부범위를 설정하였으며, 각 인자의 세부범위별 가중치를 산정하였다. 분석결과를 바탕으로 현장에서의 사용자의 편리성 및 측정 유의도를 고려하여 3-5개 범주로 구분하여 평가점수를 부여하였다. 상기과정에 의하여 산정된 각 요소별 산사태 취약성 점수를 평가표로 나타내었다.

한편, 본 연구를 기반으로 한 향후에는 지질특성별

양질의 자료에 기반한 주요 요인의 선정과 범위 설정 등의 개선 노력이 필요하다. 또한, 급경사지에서의 개발이 빈번한 우리나라에서 지질특성별 또는 특정 지역별 세부모형의 개발이 필요하며, GIS와의 연계를 통한 분석이 요구된다.

## 사 사

본 연구는 2009년도 한국연구재단의 한중일 과학기술협력확충사업의 지원으로 수행되었으며 이에 감사드립니다.

## 참고문헌

- Chae, B.-G., Kim, W.Y., Cho, Y.C., Kim, K.S., Lee, C.O. and Choi, Y.S. (2004a) Development of a logistic regression model for probabilistic prediction of debris flow. *The Journal of Engineering Geology*, v.14, p.211-222. (in Korean)
- Chae, B.-G., Kim, W.Y., Na, J.H., Cho, Y.C., Kim, K.S. and Lee, C.O. (2004b) A prediction model of landslides in the Tertiary sedimentary rocks and volcanic rocks area. *The Journal of Engineering Geology*, v.14, p.443-450. (in Korean)
- Dai, F.C., Lee, C.F. and Xu, Z.W. (2001) Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environmental Geology*, v.40, p.381-391.
- Davis, J.C., Chung C.-J. and Ohlmacher, G.C., (2006) Two models for evaluating landslide hazards. *Computers & Geosciences*, v.32, p.1120-1127.
- Gomez, H. and Kavzoglu, T. (2005) Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. *Engineering Geology*, v.78, p.11-27.
- KIGAM (2008) Development of practical technologies for damage assessments and countermeasures for geohazards. Ministry of Knowledge Economy, Korea, 486p.
- Lee, S., Ryu, J.H., Lee, M.J. and Won, J.S. (2003) Landslide susceptibility using artificial neural network at Boun, Korea. *Environmental Geology*, v.44, p.820-833.
- Lee, S., Ryu, J.H., Won, J.S. and Park, H.J. (2004) Determination and application of the weights for landslide susceptibility mapping: using an artificial neural network. *Engineering Geology*, v.71, p.289-302.
- Nandi, A. and Shakoor, A. (2009) A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Engineering Geology*,

- v.110, p.11-20.
- Kavzoglu, T. (2001) An investigation of the design and use of feed-forward artificial neural networks in the classification of remotely sensed images. Ph.D. Dissertation, University of Nottingham, School of Geography, UK, p.306.
- Van Den Eeckhaut, M., Vanwallegem, T., Poesen, J., Govers, G. Verstraeten, G, and Vandekerckhove, V. (2006) Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium). *Geomorphology*, v.76, p.392-410.
- Yilmaz, I. (2009) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey). *Computers & Geosciences*, v.35, p.1125-1138.

---

2010년 7월 6일 원고접수, 2010년 7월 22일 게재승인