

# 의사결정나무를 이용한 토양유기탄소 추정 모델 제작 Building a Model for Estimate the Soil Organic Carbon Using Decision Tree Algorithm

유수홍\* · 허준\*\* · 정재훈\*\*\* · 한수희\*\*\*\*

Yoo, Su Hong · Heo, Joon · Jung, Jae Hoon · Han, Su Hee

## 要 旨

토양유기탄소는 산림의 형성에 도움을 주며, 대기 중의 이산화탄소양을 조절함으로써 지구 온난화에 영향을 미치는 중요한 인자 중 하나이다. 토양에 존재하는 인자의 분포를 정확히 파악하려면 모든 지역에 대해 샘플링을 수행해야하나 이는 매우 비현실적인 방법이다. 따라서 알맞은 모델을 제작하여 토양유기탄소의 분포를 추정할 수 있다면 그 활용도가 높을 것으로 판단된다. 본 연구에서는 의사결정나무 알고리즘을 이용해 경사 데이터, 경사향 데이터, Digital Elevation Model (DEM), 식생의 형태 데이터로부터 토양유기탄소를 상대적으로 다량 함유하고 있는 환경 인자를 파악할 수 있는 모델을 제작했으며, 정확도 검증은 10 집단 교차 검정을 통해 수행하였다. 이를 위하여 See 5와 Weka 소프트웨어를 이용하였다. See5 소프트웨어의 경우, 토양유기탄소 표층에 대해 식생의 형태에 의해 토양유기탄소량이 결정되는 것으로 나타났으며, 중간층에 대해서는 DEM에 의해 토양유기탄소량이 달라진다는 결론이 도출됐다. 생성된 모델의 정확도는 표층에 대해 70.8%, 중간층에 대해 64.7%인 것으로 나타났다. Weka 소프트웨어의 경우, 토양유기탄소 샘플의 표층에 대해 See5와 동일한 결과가 도출되었지만, 중간층에 대해서는 DEM이나 식생의 형태뿐만 아니라 경사향도 영향을 미친다는 결론이 도출되었다. 생성된 모델의 정확도는 표층에 대해 68.98%, 중간층에 대해 60.65%인 것으로 나타났다. 본 연구는 토양유기탄소량의 파악 및 토양유기탄소 지도 제작에 활용될 수 있을 것으로 사료된다.

핵심용어 : 토양유기탄소, 의사결정나무 알고리즘

## Abstract

Soil organic carbon (SOC), being a help to forest formation and control of carbon dioxide in the air, is found to be an important factor by which global warming is influenced. Excavating the samples by whole area is very inefficient method to discovering the distribution of SOC. So, the development of suitable model for expecting the relative amount of the SOC makes better use of expecting the SOC. In the present study, a model based on a decision tree algorithm is introduced to estimate the amount of SOC along with accessing influencing factors such as altitude, aspect, slope and type of trees. The model was applied to a real site and validated by 10-fold cross validation using two softwares, See 5 and Weka. From the results given by See 5, it can be concluded that the amount of SOC in surface layers is highly related to the type of trees, while it is, in middle depth layers, dominated by both type of trees and altitude. The estimation accuracy was rated as 70.8% in surface layers and 64.7% in middle depth layers. A similar result was, in surface layers, given by Weka, but aspect was, in middle depth layers, found to be a meaningful factor along with types of trees and altitude. The estimation accuracy was rated as 68.87% and 60.65% in surface and middle depth layers. The introduced model is, from the tests, conceived to be useful to estimation of SOC amount and its application to SOC map production for wide areas.

Keywords : SOC(Soil Organic Carbon), Decision Tree Algorithm

2010년 7월 21일 접수, 2010년 8월 24일 채택

\* 정회원 · 연세대학교 공과대학 사회환경시스템공학부 석사과정(swennoir@yonsei.ac.kr)

\*\* 정회원 · 연세대학교 공과대학 사회환경시스템공학부 부교수(jheo@yonsei.ac.kr)

\*\*\* 연세대학교 공과대학 사회환경시스템공학부 박사과정(lionheart\_kr@yonsei.ac.kr)

\*\*\*\* 교신저자 · 연세대학교 공과대학 사회환경시스템공학부 박사 후 과정(scivile@yonsei.ac.kr)

## 1. 서 론

토양유기탄소는 식물의 광합성 작용에 의해 고정된 탄소가 생물의 잔해로 토양에 들어온 후 토양 내에서 분해되는 과정 중에 있는 물질로써, 토양 속에서 굉장히 느린 속도로 분해되면서 토양을 비옥하게 하거나 대기 중에 분포하고 있는 이산화탄소와의 관계를 조절하는 중요한 요소이다. 따라서 토양유기탄소의 분포를 파악할 수 있는 모델을 제작한다면 산림의 형성 정책 결정 및 개발부지 선정시 중요한 참고자료로서 사용될 수 있다.

토양유기탄소의 분포를 파악하기 위하여 Chen, F. 등(2000)이 그루지아의 Crisp 주에 대해 토양유기탄소 모델을 제작하였다. 대상 지역에 대해 샘플링한 토양유기탄소의 함유량과 원격 탐측을 통해 취득한 영상의 RGB 밴드간에 알맞은 로그 선형 식(logarithm linear equation)을 생성하여 77.4%의 정분류율(classified rate)을 가지는 모델을 제작했다. 그러나 해당 모델은 원격 탐측을 통해 수집한 정보에 의존하므로 영상의 밝기값에 따라 전혀 다른 결과가 도출될 가능성이 있다.

의사결정나무 알고리즘은 데이터마이닝 기법 중 하나로써 알고리즘에 의해 생성된 규칙에 따라 모든 데이터가 하나의 클래스(Class)에 속할 수 있도록 모델을 형성하는 알고리즘을 말한다(이극노, 2003). 다양한 분야에 적용될 수 있는 의사결정나무 알고리즘은 편마암 지역에서의 급경사지 재해 예측 시스템 개발에 이용된 바 있다. 송영석 등(2008)이 수행한 연구에서는 지형고도, 사면경사, 포화도, 투수계수, 산사태 발생 위치, 토양의 종류 등의 다양한 데이터들을 의사결정나무 알고리즘에 적용하여 사면경사가 17.9°, 포화도가 52.1%, 사면고도가 320m에 해당하는 경우, 급경사지재해가 발생할 확률이 높은 것으로 추정했다. 또한 유역 특성에 맞게 홍수방어시설물을 조합함으로써 유기적인 방어 기능을 도모하기 위해 이용된 바 있다. 변성호 등(2008)은 의사결정나무 알고리즘을 이용해 특정 요구에 알맞은 홍수방어시설물 선정 모델을 제작했으며, 이를 이용해 문산천 홍수방어대책을 도출하고 유역에 따라 사전에 결정된 홍수방어대안과의 비교를 수행했다. 이외에도 의사결정나무는 이동통신고객 분류(이극노, 2003), 산업 재해 예측 모형 제작(임영문, 2005), 토양 분포 예측(박수진, 2010) 등 다양한 분야에서 특정 현상에 대한 예측 및 분류를 위해 이용되고 있다.

본 연구에서는 의사결정나무를 제작하기 위해 토양유기탄소량과 네 가지 인자(경사, 경사향, 식생의 형태, DEM)를 이용하였다. 토양유기탄소는 표층, 중간층, 바

닥층에 대해 추출되었으며, 각 층에 대한 함유량이 GIS의 점 데이터 형태로 저장되어 있다. 이 데이터를 내추럴 브레이크(Natural Break) 방법에 의해 각각의 클래스(Class)로 구분하여 의사결정나무 제작에 이용하였다.

## 2. 실험 대상 지역 및 입력 자료의 종류

본 연구에서 선정한 연구 대상지는 인도의 Kolli hill 지역으로, 지역의 중심에는 도시가 존재하고 있으며 그 주변으로 산림이 분포하고 있는 특이 지형으로 활엽수와 침엽수 그리고 재배지가 분포하고 있다. 위치는 북위 11°10'00"~11°30'00", 동경 78°15'00"~78°30'00"이며 면적은 약 503km<sup>2</sup>이다(그림 1).

본 연구에서 이용한 데이터는 토양유기탄소 함유량, Digital Elevation Model(DEM)과 경사 데이터, 경사향 데이터, 서식하고 있는 식생의 형태이다. 토양유기탄소의 데이터를 얻기 위해 대상 지역에 대해 채취한 샘플링 데이터를 이용했다. 이 데이터는 GIS 형태의 점 데이터로 되어 있으며 2km<sup>2</sup>에 대해 한 번의 샘플링이 수행되었다. 총 216개의 샘플 토양유기탄소의 샘플은 표면의 일들을 제거한 다음, 1m×1m의 면적으로 채취하되 깊이에 따라 표층(0~30cm), 중간층(30~60cm), 바닥층(60~90cm)으로 구분했다. 샘플은 실험실에서 건조시켜 각각의 층에 대해 2mm 체에 통과시켜 투과되지 않은 샘플만을 취득했으며, 각 층에 대해 clad 방법(Ramachandran, A., 2007)을 통해 부피 밀도(mg·m<sup>-3</sup>)를 계산했다. 토양유기탄소의 밀도(mg·ha<sup>-1</sup>)는 앞서 계산된 부피 밀도를 percent coarse fraction에 대해 보정함으로써 추정하였다(식 (1), (2)).

$$\text{조정된 부피 밀도}(mg \cdot m^{-3}) = \text{부피 밀도}(mg \cdot m^{-3}) \quad (1)$$

$$\times \frac{(100 - \text{percent coarse fraction})}{100}$$

$$\text{토양유기탄소 밀도}(mg \cdot ha^{-1}) = \frac{\text{토양유기탄소}(\%)}{100} \quad (2)$$

$$\times \rho_b \times \text{깊이}(m) \times 10^4 (m^2 \cdot ha^{-1})$$

단,  $\rho_b$ : 조정된 부피 밀도

DEM은 Kolli hill 지역의 등고선 수치 지도를 이용하여 TIN을 생성한 다음 이를 래스터 형태로 변환하여 제작하였다. 경사 데이터는 지형의 기울기(gradient)를 x와 y방향에 대해 나타낸 편미분 요소의 백터 크기로 나타냄으로서 데이터를 제작하였다(Hunter, Gary J., 1997). 경사향 데이터는 x와 y방향에 대해 나타낸 편미

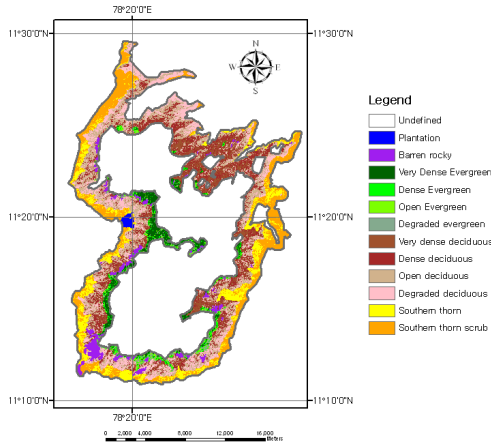


그림 1. 실험 대상지 - Kollil hill

분 요소를 나눔으로써 제작하였다. 본 연구에서는 상용 소프트웨어인 ArcGIS를 이용하여 DEM과 경사 데이터, 경사향 데이터를 제작하였는데, 이에 사용된 알고리즘은 앞서 언급한 알고리즘과 동일하다. 식생 형태 데이터는 대상 지역에 대한 현장 조사를 통해 취득된 데이터로서 polygon 형식으로 저장되어 있다. 한편 조사된 식생의 종류에 대한 정보를 표 1에 나타내었으며, 그 외 본 연구에서 이용한 데이터들이 포함하고 있는 정보와 그에 대한 설명을 간략히 나열하였다.

표 1. 연구에 이용된 데이터

데이터 종류	데이터 정보
토양유기탄소 샘플	- GIS 형식의 점 데이터 (총 216개)
DEM	- 160m와 1415m 사이의 표고를 가짐
경사	- 도(°) 단위로 표현되어 있으며 0°에서 88° 사이의 값을 가짐
경사향	- 도(°) 단위로 표현되어 있으며 0°에서 360° 사이의 값을 가짐 - 0° ~ 22.5° & 337.5° ~ 360° : 북, 67° ~ 112.5° : 동, 157.5° ~ 202.5° : 서, 247.5° ~ 292.5° : 남 - 평평한 지역에 대해서는 -1의 값을 가짐
식생의 형태	- (a) 재배지와 (b) 암석지, (c) 매우 조밀히 분포된 상록수 분포 지역, (d) 조밀히 분포된 상록수 분포 지역, (e) 드문드문 분포된 상록수 분포 지역, (f) 퇴화된 상록수 분포 지역, (g) 매우 조밀히 분포된 낙엽수 분포 지역, (h) 조밀히 분포된 낙엽수 분포 지역, (i) 드문드문 분포된 낙엽수 분포 지역, (j) 퇴화된 낙엽수 분포 지역, (k) 남향 관목 분포 지역, (l) 남향 관목 분포 지역

### 3. 모델의 구축

#### 3.1 의사결정나무(Decision Tree)

의사결정나무는 의사결정규칙을 도표화하여 관심대상이 되는 집단을 몇 개의 소집단으로 분류하거나 예측을 수행하는 계량적 분석방법을 말한다. 분석결과는 ‘조건이 A이고 조건이 B이면 결과집단 C’라는 형태의 규칙으로 표현되므로 분류 또는 예측을 목적으로 하는 다른 계량적 분석방법에 비해 쉽게 이해할 수 있다는 장점이 있다.

의사결정나무의 구성은 노드(node)로 되어 있으며, 뿌리 노드(root node)로부터 시작하여 분리기준, 정지규칙, 가지치기 등에 의해 각 가지가 노드의 끝에 이를 때까지 노드를 형성함으로써 가장 하위 노드인 잎 노드(leaf node)까지 만들어진다. 이 중 분리기준은 데이터셋을 효율적으로 나눌 수 있는 기준을 의미하며, 가지치기는 오분류율(misclassified rate)을 발생시킬 수 있는 부적절한 규칙을 제거하는 작업을 말한다.

본 연구에서 이용한 C4.5 알고리즘은 J. Ross quinlan에 의해 개발된 알고리즘으로서(임영문, 2005), 각 하위 노드들에 대한 엔트로피를 계산한 뒤 이 엔트로피가 높은 상태에서 낮은 상태가 되도록 특정 조건을 찾아 나무 모양으로 구분해 나가는 방식이다(The University of Trier). 주어진 데이터 집합의 혼잡도를 의미하는 엔트로피는 주어진 데이터 집합에 얼마나 다양한 레코드들이 존재하는가에 대한 정보를 지수로 나타낸 것으로서 같은 종류의 레코드들이 많이 존재하면 엔트로피 지수가 낮고, class의 종류가 다양하면 높은 엔트로피 지수를 가진다. 식 (3)은 데이터셋 S에 대해 엔트로피를 계산하는 식으로 각 class 값의 포함 비율에 로그를 적용한 뒤, 각각의 가중치로 곱한 값을 모두 더한 것으로 나타낸다.

$$Entropy(S) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (3)$$

$$\text{단, } p_i = \frac{freq(C_i, S)}{|S|}$$

S : 주어진 데이터들의 집합,

C = C<sub>1</sub>, C<sub>2</sub>, ..., C<sub>k</sub> : class 값들의 집합,

freq(C<sub>i</sub>, S) : S에서 class C<sub>i</sub>에 속하는 레코드의 수, |S| : 주어진 데이터들의 집합의 데이터 개수

데이터 셋의 엔트로피를 계산하고 나면 정보 이득(information gain) 과정을 수행하게 되는데, 이는 어떠한

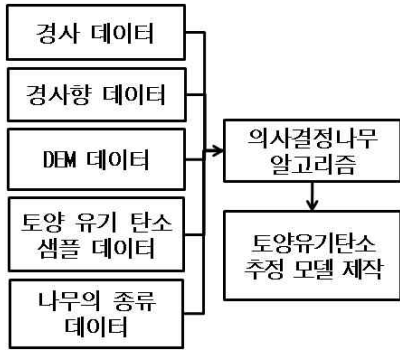


그림 2. 연구 흐름도

한 속성을 선택했을 때 전체 데이터를 더 잘 구분시킬 수 있는가에 대한 정보를 나타낸다. 식 (4)는 속성 A를 선택했을 때의 정보 이득 양을 계산하는 식으로써, 이 값이 클수록 정보 이득이 크다는 것을 의미하게 되고 이는 속성 A의 변별력이 좋다는 것을 의미한다.

$$\text{Gain}(A) = I(S_1, S_2, \dots, S_m) - E(A) \quad (4)$$

단,  $\text{Gain}(A)$ : 속성 A의 이득 양,  
 $I(S_1, S_2, \dots, S_m)$ : 상위 노드의 엔트로피 값,  
 $E(A)$ : A 속성을 선택했을 때, m개의 하위 노드들에 대한 엔트로피의 평균값

구축된 모델의 평가 방법 중 하나인 정오분류표는 제작된 모델을 이용해 분류를 수행했을 때 올바르게 분류된 속성과 그렇지 않은 속성의 수를 하나의 표로 나타내어 모델의 정확도를 파악할 수 있다. 또한 구축된 모델의 분석용 자료(training data)와 평가용 자료(validation data)의 정분류율과 평가용 자료의 오분류율을 산정하여 제작된 모델의 성능 평가를 수행했다. 본 연구에서는 각각 C5.0과 C4.5 알고리즘이 적용된 상용 소프트웨어 See5(RuleQuest Research)와 Weka(The University of Waikato)를 이용하여 의사결정나무 모델을 구축했다.

C5.0 알고리즘은 C4.5 알고리즘과 거의 유사한 알고리즘이나 메모리 사용의 효율성을 높이고, 처리 속도를 향상시킨 알고리즘을 말한다.

### 3.2 종속변수 데이터 생성

본 연구에서는 내추럴 브레이크 분류 방법을 통해 토양유기탄소 샘플 데이터를 분류하여 종속변수 자료로 사용했다. 내추럴 브레이크 분류 방법은 저장된 데이터 내에서 상대적으로 값이 급증하는 단계에서 데이터를 구분시키는 방법이다. 샘플링 데이터를 표층, 중간층,

표 2. 구분된 토양유기탄소량

층	Class	함유량(탄소의 중량(g) / 흙의 중량(kg))
표층	class1	0.06~1.28
	class2	1.28~3.34
	class3	3.34~9.77
중간층	class1	0.07~0.76
	class2	0.76~2.27
	class3	2.27~5.38
바닥층	class1	0.06~0.98
	class2	0.98~2.70
	class3	2.70~4.92

바닥층에 대해 총 3개의 클래스 즉, class1, class2, class3로 구분하였다. Class 1의 경우 가장 적은 양의 탄소를 저장하고 있는 샘플 데이터들의 집합이며, class 3에 가까이 분류될수록 많은 양의 탄소를 저장하고 있는 샘플들로 구성했다. 표 2는 표층, 중간층, 바닥층에 대해 각각 분류된 클래스가 포함하고 있는 토양유기탄소의 함유량을 나타낸 것이다.

### 4. 알고리즘의 적용 및 정확도 분석

See5 소프트웨어를 통해 도출된 결과는 그림 3과 같으며, 뿌리 노드에서 잎 노드까지 텍스트 기반으로 나타났다. 각 잎 노드의 끝에는 (n) 또는 (n/m)이 표기되어 있는데, n은 해당 잎 노드와 동일한 종속변수를 가진 속성의 수를 나타내고, m은 그렇지 못한 속성의 수를 나타낸다.

토양유기탄소 표층의 경우, 매우 조밀히 분포되어 있는 상록수 지역일 때 상대적으로 더 많은 양의 토양유

```
Decision tree:
Vegetation in {a,e,f}: Class2 (15.3/6.3)
Vegetation in {b,d,g,h,i,j,k,l}: Class1 (178.3/49)
Vegetation = c: Class3 (22.4/6.4)
```

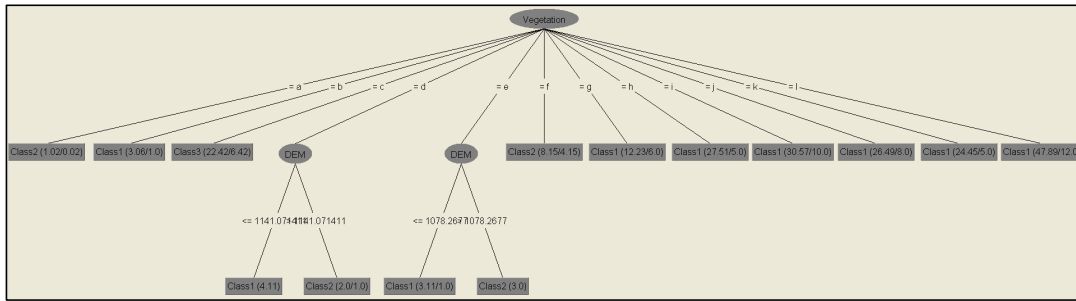
(a)

```
Decision tree:
DEM <= 1003.513:
... DEM <= 631.45: Class1 (123/33)
... DEM > 631.45: Class2 (51/15)
DEM > 1003.513:
... Vegetation in {a,g,h,i,j,k,l}: Class3 (0)
... Vegetation in {b,e}: Class2 (7/2)
... Vegetation = c: Class3 (21/6)
... Vegetation in {d,f}: Class1 (14/7)
```

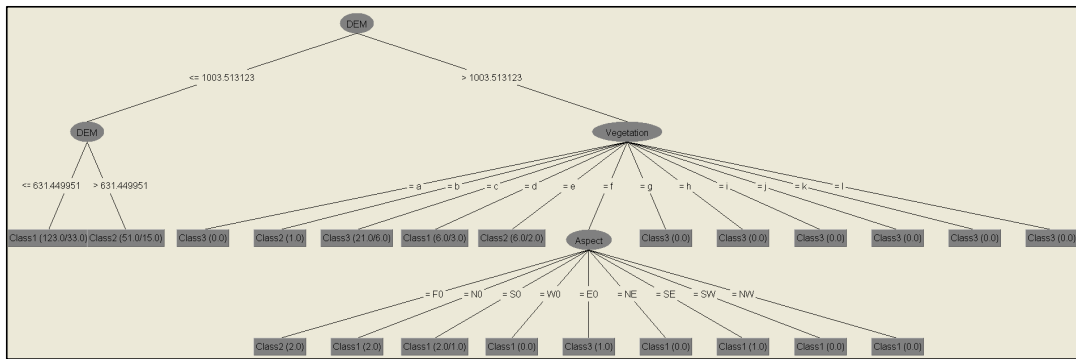
(b)

그림 3. See5 소프트웨어를 통해 도출된 결과 -

(a) 표층, (b) 중간층



(a)



(b)

그림 4. Weka 소프트웨어를 통해 도출된 결과 - (a) 표층, (b) 중간층

기탄소를 함유한다는 결론이 도출되었다. 중간층에 대해서는 표고가 1003.513m 이상이고, 매우 조밀히 분포한 상록수 지역일 때 상대적으로 더 많은 양의 토양유기탄소를 함유한다는 결론이 도출되었다. 바닥층에 대해서는 모델이 형성되지 않았다.

Weka 소프트웨어를 통해 도출된 결과는 그림 4와 같으며, 토양유기탄소 샘플의 표층에 대해 매우 조밀히 분포된 상록수지역이 상대적으로 많은 양의 토양유기탄소를 가지는 것으로 도출됐다. 토양유기탄소 샘플의 중간층에 대해서는 표고가 1003.513m 이상이고, 서식하고 있는 식생의 형태가 매우 조밀히 분포한 상록수 지역이거나 경사향이 동쪽인 지역에 분포한 퇴화된 상록수 지역일 때 상대적으로 많은 양의 토양유기탄소를 가지는 것으로 도출되었다. 바닥층에 대해서는 See 5와 마찬가지로 모델이 형성되지 않았다.

두 소프트웨어 모두 표층에 대해 식생의 형태가 최상위 분리기준변수로 선정되었는데, 이는 낙엽과 같은 바이오매스와의 직접적인 접촉 기회가 많기 때문인 것으로 판단된다. 중간층에 대해서는 DEM이 선정되었는데, 연구 대상지의 경우 고도가 높은 지역에 대개 활엽수가 분포하고 있기 때문으로 판단된다. 반면, 바닥층에 대해서는 본 연구에서 이용한 환경 인자 중, 그 어떤

인자도 토양유기탄소량에 영향을 미치지 않는 것으로 나타났는데, 이는 지상과의 접촉이 거의 없기 때문인 것으로 판단된다.

제작된 의사결정나무의 정확도는 올바르게 분류된 속성 수를 전체 속성 수로 나눈 정분류율을 계산함으로써 모델의 정확도를 파악할 수 있다. 형성된 모델의 타당성은 10 집단 교차 검증(10 fold cross validation)을 통해 평가했다. 표 3과 표 4은 See5 소프트웨어에서 도출된 토양유기탄소의 표층과 중간층에 대한 정오분류표이고, 표 5와 표 6은 Weka 소프트웨어에서 도출된 토양유기탄소의 표층과 중간층에 대한 정오분류표이다.

표 3. 토양유기탄소 샘플의 표층에 대한 정오분류표 (See 5)

분류된 값 \ 실제 관측된 값	Class 1	Class 2	Class 3	합계
Class 1	129	5	1	135
Class 2	45	8	5	58
Class 3	5	2	16	23
합계	179	15	22	216

$$- \text{정분류율} = \frac{129 + 8 + 16}{216} \times 100 = 70.8\%$$

표 4. 토양유기탄소 샘플의 중간층에 대한 정오분류표 (See 5)

분류된 값 \ 실제 관측된 값	Class 1	Class 2	Class 3	합계
Class 1	90	24	2	116
Class 2	42	36	5	83
Class 3	2	1	14	17
합계	134	61	21	216

- 정분류율 =  $\frac{90 + 36 + 14}{216} \times 100 = 64.7\%$

표 5. 토양유기탄소 샘플의 표층에 대한 정오분류표 (Weka)

분류된 값 \ 실제 관측된 값	Class 1	Class 2	Class 3	합계
Class 1	128	6	1	135
Class 2	48	5	5	58
Class 3	5	2	16	23
합계	181	13	22	216

- 정분류율 =  $\frac{129 + 5 + 16}{216} \times 100 = 68.98\%$

표 6. 토양유기탄소 샘플의 중간층에 대한 정오분류표 (Weka)

분류된 값 \ 실제 관측된 값	Class 1	Class 2	Class 3	합계
Class 1	86	28	2	116
Class 2	45	31	7	83
Class 3	2	1	14	17
합계	133	60	23	216

- 정분류율 =  $\frac{86 + 31 + 14}{216} \times 100 = 60.65\%$

See 5의 경우, 토양유기탄소 샘플의 표층에 대해 70.8%, 중간층에 대해 64.7%의 정분류율을 가지는 것으로 나타났으며, Weka 소프트웨어의 경우, 표층에 대해 68.98%, 중간층에 대해 60.65%의 정분류율을 가지는 것으로 나타났다.

### 5. 결 론

본 연구에서는 의사결정나무를 이용하여 토양유기탄소량 추정 모델을 제작하였다. 토양유기탄소량에 영향

을 미칠 수 있는 인자로서 DEM, 경사, 경사향, 식생의 형태를 고려하였으며, 의사결정나무를 구현하기 위해 상용 소프트웨어인 See5와 Weka 소프트웨어를 이용하였다. 두 소프트웨어 모두 표층에 대해서는 식생의 형태가 가장 중요한 환경 인자인 것으로 나타났다. 중간층에 대해서는 DEM이 가장 중요한 인자로 도출되었다. 이는 연구 대상지의 지형 특성상 고도가 높은 지역에 상록수와 같은 활엽수가 분포하고 있으므로, 결론적으로는 식생의 형태가 토양유기탄소량에 영향을 미치는 것으로 판단된다. 바닥층에 대해서는 두 소프트웨어 모두 의사결정나무가 형성되지 않았는데, 이는 지상의 자연적 환경이 바닥층에 대해 영향을 미치지 않기 때문인 것으로 판단된다. 생성된 의사결정나무 모델은 토양유기탄소량의 예측 및 지도 제작에 활용될 수 있을 것으로 사료된다.

본 연구에서는 토양유기탄소량을 구분하기 위해 내추럴 브레이크 방법을 이용하였으나, 구분 방법에 따라 다른 모델이 생성될 것으로 판단된다. 따라서 이에 대한 연구가 필요할 것으로 사료된다.

### 참고문헌

1. 박수진, 손연규, 홍석영, 박찬원, 장용선, 2010, "한국 주요 토양유형의 공간적 분포와 토양형성요인을 이용한 예측가능성 평가", 대한지리학회지, 제 45권, 제 1호, pp.95-118.
2. 변성호, 강현직, 한정우, 김태웅, 2008, "의사결정나무모형을 이용한 유역내 구조적 홍수방어 대안 도출", 대한토목학회지, 제 28권 제 1 B 호, pp.33-40.
3. 송영석, 채병곤, 2008, "의사결정나무모형을 이용한 편마암 지역에서의 급경사지재해 예측기법 개발", The Journal of Engineering Geology, Vol.18, pp.45-54.
4. 이극노, 이홍철, 2003, "이동통신고객 분류를 위한 의사결정나무와 신경망 결합 알고리즘에 관한 연구", 한국지능정보시스템학회논문지, 제9권, 제1호, pp.139-155.
5. 임영문, 박준구, 황영섭, 2005, "C4.5 알고리즘을 이용한 산업 재해의 특성 분석", 한국안전학회지, 제20권, 제4호, pp.130-137.
6. Chen, F., Kissel, D. E., West, L. T. and Adkins, W., 2000, "Field-Scale Mapping of Surface Soil Organic Carbon Using Remotely Sensed Imagery", Soil Sci Soc America, Vol.64, pp.746-753.
7. Hunter, Gary J.와 Goodchild, Michael F., 1997, "Modeling the Uncertainty of Slope and Aspect Estimates Derived from Spatial Databases", Geographical Analysis, Vol.29, No.1, pp.35-49.

8. Ramachandran, A., Jayakumar, S., Haroon, R. M., Bhaskaran, A. and Arockiasamy, D. I., 2007, "Carbon sequestration: estimation of carbon stock in natural forests using geospatial technology in the Eastern Ghats of Tamil Nadu", India, Current Science, Vol.92, pp.323-331.
9. RuleQuest Research, Australia, <http://www.rulequest.com/see5-info.html>
10. Scull P., Franklin, J., Chadwick O.A. and McArthur, D., 2003, "Predictive soil mapping - a review", Progress in Physical Geography, Vol.27, pp.171-197.
11. Skidmore, A.K., Watford, F., Luckananurug, P. and Ryan, P.J., 1996, "An operational GIS expert system for mapping forest soi",. Photogrammetric Engineering and Remote Sensing, Vol.62, pp.501-511.
12. The University of Trier, Howard J. Hamilton, <http://www2.cs.uregina.ca/~dbd/cs831/course.html>
13. The University of Waikato, New Zealand, <http://www.cs.waikato.ac.nz/~ml/index.html>