

# 위성영상과 데이터 마이닝 기법을 이용한 토양수분 산정

## Estimate Soil Moisture Using Satellite Image and Data Mining

김광섭\*, 박한균\*\*, 조소현\*\*\*

Kim, Gwangseob, Park, Han Gyun, Cho, So Hyun

### 요지

토양수분은 토양입자에 포함되어 있는 물을 의미하는 것으로 지표면과 대기간의 에너지 균형과 물 순환을 조절하는데 중요한 요소이다. 본 연구에서는 토양수분 산정을 위하여 2003년 1월부터 2008년 12월까지의 MODIS(Moderate Resolution Imaging Spectroradiometer) 위성관측 자료로부터 획득한 정규식생지수(NDVI : Normalized Difference Vegetation Index)자료와 지표면 온도자료, 우리나라 76개소 기상관측소 중에 자료의 보유기간이 30년 이하인 관측소와 섬 지역들을 제외한 57개 지점의 강수량, 토양온도 자료 및 우리나라 전역에 대한 토지피복, 유효토심자료를 이용하여 데이터 마이닝(Data Mining) 기법의 하나인 CART(Classification And Regression Tree) 기법을 이용하여 토양수분을 산정하였다. 먼저 신뢰성 높은 토양수분 관측 자료를 가진 용담댐 유역의 6개 지점에 대하여 토양수분을 산정하여 적용 가능성을 분석하였다. 3개 지점의 토양수분 관측치는 토양수분 산정 모형 수립에 사용하였으며 검증에 사용된 1개 지점의 토양수분의 관측치와 추정치 간의 상관계수를 확인한 결과 전체적인 토양수분의 거동을 잘 나타내고 있어 토양수분 추정 모형의 적용가능성을 확인하였다. 이를 이용하여 용담댐 유역의 토양수분 분포와 우리나라 전역에 대한 토양수분 분포도를 추정하였다. 신뢰할 수 있는 지상관측 토양수분 관측치가 다양한 지상조건에 대하여 존재하지 않는 한계가 있음에도 불구하고 제시된 토양수분산정 방법은 제한된 가용자료를 사용한 우리나라 전역의 토양수분 산정에 있어 합리적인 접근법이라 판단된다.

**핵심용어 :** 토양수분, 데이터 마이닝, 의사결정나무

## 1. 서 론

토양수분의 시공간적인 특성은 증발, 침투, 지하수, 식생 등에 영향을 미치는 중요한 변수이다. 토양수분을 측정하는 방법에는 세타 탐침(Theta probe), 장력계, TDR(Time Domain Reflectometry)등이 이용되고 있으며(이가영 외, 2004), 광역에 대한 토양수분자료의 보다 정확한 공간변동성을 관측하기 위하여 항공원격탐사(airborne remote sensing)와 인공위성 원격탐사(spaceborne remote sensing)기술이 개발되어 적용되고 있으나 국내에서는 항공원격탐사를 위한 장비의 부재로 인하여 관측경험이 전무한 실정이며, 토양수분에 대한 인공위성 원격탐사 자료는 NASA의 지구관측 시스템에서 제공하는 Advanced Microwave Scanning Radiometer - Earth Observing System(AMSR-E) 자료를 사용하고 있는 실정이다. 본 연구에서는 MODIS(Moderate Resolution Imaging Spectroradiometer) 위성관측 자료로부터 획득한 정규식생지수(NDVI : Normalized Difference Vegetation Index)자료와 지표면 온도자료, 용담댐유역에 설치된 4개 지점(부귀, 상진, 천천2, 계북2)의 강수량자료, 토양온도 자료 및 토지피복, 유효토심자료를 이용하여 데이터 마이닝(Data Mining) 기법의 하나인 CART(Classification And Regression Tree) 기법을 이용하여 토양수분을 산정하였다.

\* 정회원 · 경북대학교 건축 · 토목공학부 토목공학전공 교수 · E-mail : kimg@knu.ac.kr

\*\* 경북대학교 건축 · 토목공학부 토목공학전공 석사과정 · E-mail : harusmile@knu.ac.kr

\*\*\* 경북대학교 건축 · 토목공학부 토목공학전공 석사과정 · E-mail : shcho35@knu.ac.kr

## 2. 연구방법

본 연구에서는 여러 변수를 이용하여 목표변수를 산정하는 규칙들의 집합을 생성하는데 준비된 자료를 일정한 분리기준에 따라 이진분리를 실시하여 동질 그룹을 분리하여 의사결정나무를 구성하는 CART기법을 이용한다. 또한 모형의 타당성을 평가하기 위하여 훈련자료(Training data)와 검증자료(Test data)로 분리 적용하는 교차검증 방식인 10배 교차 검증법(K-fold cross validation method)을 사용하여 신뢰성을 높였다.

우리나라의 토양수분 산정에 앞서 먼저 용담댐 유역의 지상관측망 토양수분 자료와 지형, 토지피복, 식생지수 및 강수량 자료 등을 이용하여 토양수분간의 관계를 파악하였고 또한 식생의 변화를 고려한 토양수분 추정을 위하여 MODIS NDVI자료를 이용하였다. 수집된 토양수분 자료, 위성영상 자료, 지형자료 등을 이용하여 토양수분의 변동성을 분석하여 토양수분을 산정할 수 있는 데이터마이닝(Data mining)기법을 선택하였다. 널리 사용되는 의사결정나무기법 중에 하나인 CART 기법을 이용하여 자료의 상관성을 분석하고 하나의 의사결정나무(Decision Tree)를 완성하여 우리나라 전역의 토양수분 산정의 적용 가능성을 확인하였다.

## 3. 연구결과

### 3.1 대상 유역

본 연구에서는 토양수분을 산정하기 위하여 용담댐의 토양수분 자료를 이용하고자 한다. 용담댐유역은 수자원관리정보시스템(WAMIS)에서 획득한 토양도(유효토심, 종류)에 따르면, 면적의 62.09%가 얇은 토심을 이루고 있으며 토양의 종류는 면적의 65.33%가 사양질과 식양질, 9.76%가 식질로 이루고 있다. 산림의 41.48%가 활엽수림으로 이루어져 있으며, 침엽수림은 36.20%의 비율을 차지하고 있다. Fig. 1은 대상유역인 용담댐 유역과 우량관측소의 위치와 각각 유효토심별, 종류별 토양도, NDVI, 토양온도, 강수량 자료 샘플을 나타낸 그림이다.

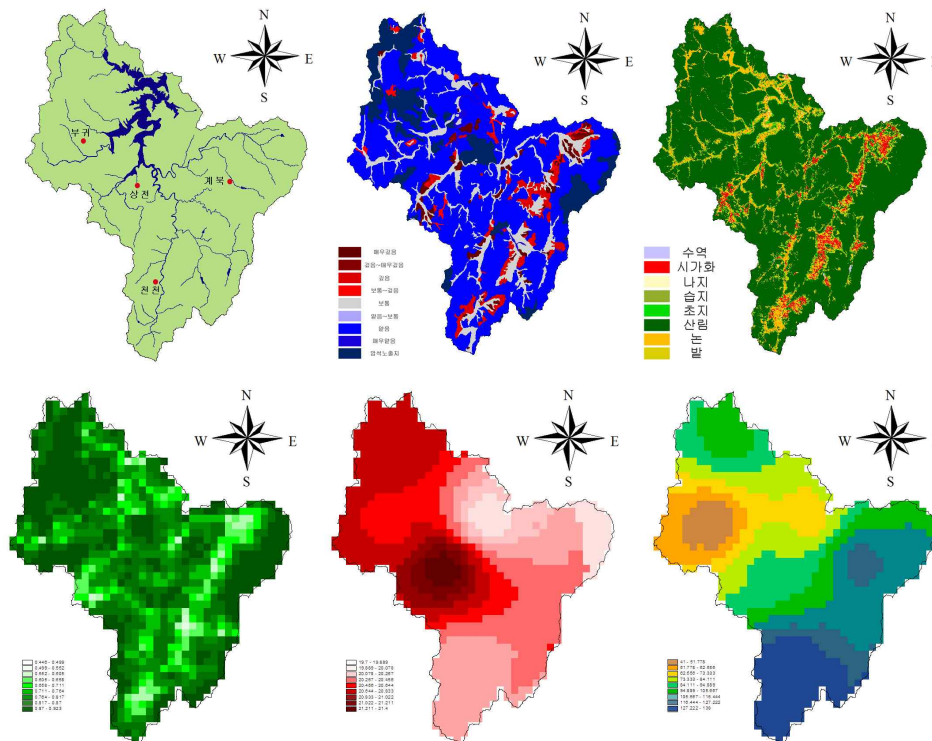


Fig.1. 용담댐 유역의 우량관측소와 자료 샘플

토양수분을 산정하기 위하여 용담댐유역에 설치된 4개 지점(부귀, 상전, 천천2, 계북2)의 강수량자료와 토양수분자료를 이용하였으며, 자료의 기간은 2008년 5월 16일부터 8월 19일까지이고 일 단위 토양수분 자료이다. Fig. 2는 용담댐유역에 설치된 4개 지점 중 한 지점인 부귀관측지점의 토양수분과 강수량 자료의 시계열을 나타낸 그림으로, 시계열 자료가 보여주는 강우사상에 대한 토양수분의 반응과 비 강수기간동안의 토양수분 감쇄 등 전반적인 거동을 잘 나타내고 있다.

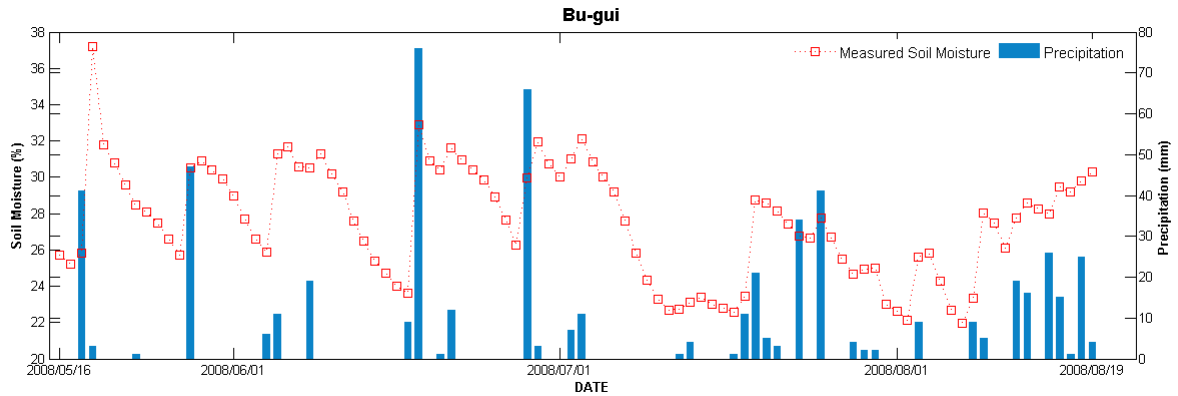


Fig.2. 부귀의 토양수분과 강수량 자료

Fig.3. 용담댐 유역의 토양 자료

본 연구에서는 토양수분 산정을 위하여 사용하는 입력 변수는 MODIS 위성관측 자료로부터 획득한 우리나라 전역에 대한 정규식생지수와 지표면온도자료를 사용하였다. 정규식생지수(NDVI)는 근적외선 파장대(Near-InfraRed)와 적색영역(Red) 반사도 값의 합과 차의 비로 계산되며, 계산식은 Eq. (1)로 표현된다.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

### 3.2 용담댐 지역의 토양수분 산정

공간적으로 향상된 토양수분자료를 획득하기 위해서 CART 분석에 사용된 변수는 대상 유역의 토지피복, 토양도(유효토심, 종류), MODIS LST, MODIS NDVI, 토양온도(°C), 강수량(mm), 지점 토양수분(%) 등 8가지의 변수가 사용되었다. Table 1은 관측지점의 변수별 기초 통계량을 나타낸 표로 NDVI의 경우에는 최소 0.4860부터 최대 0.8530의 값으로 분포되어 있으며, 대부분의 변수들의 경우에는 NDVI와 같이 비슷한 수치대로 자료가 구성되어 있으며, 토양수분 실측치의 경우에는 주천과 안천의 경우 10% 이하로 나타난 경우가 있지만 보통 최소 10%대에서 최대 38%까지 다양하게 분포되어 있다. 본 연구에 수집된 자료 중에 CART 기법에 적용한 지점은 부귀, 상전, 천천2 총 3개의 지점이고, 주천, 안천 지점은 CART 기법 분석 후 생성된 산정식의 검증자료로 사용하였다.

Station	Soil Depth	Land cover	Soil type	NDVI		Maximum Precipitation (mm/day)	Soil temperature(°C)		Soil moisture(%)	
				Min	Max		Min	Max	Min	Max
부귀	보통	산림	식양질	0.7820	0.8530	76	16	27.2	21.9	37.2
안천	얕음	산림	식양질	0.7560	0.8270	69	15.8	25.6	6.0	18.5
천천2	얕음	산림	식양질	0.6950	0.8100	138	15.3	27.6	13.1	29.1
계북2	깊음	산림	식질	0.4860	0.7600	121	16.1	26.6	9.35	38.0

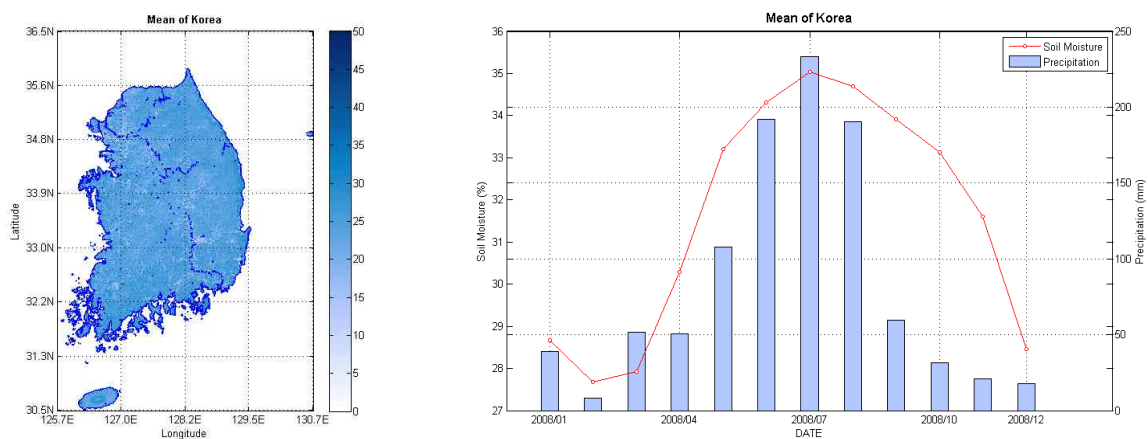
Table 1. 관측지점의 기초 통계량

대표적인 분류나무 구성법인 CART의 장점은 앞서 진술한 것과 같이 수집된 자료의 형태상 온도, 지면온도와 같이 연속형 자료와 유효토심별 토양도, 토지피복과 같이 범주형 자료를 이용하여 최적 그룹을 생성하는 규칙이 사용자가 이해하기 쉽다는 것이다. 용담댐 지역의 대상기간 동안의 자료에 적용한 결과 5개의 최적 산정식을 생성하였으며 산정식을 이용한 토양수분 추정치와 관측된 토양수분의 상관성을 분석한 결과로서 상관계수는 평균 0.747의 상관계수를 보였다. 이는 생성된 토양수분으로 적절한 추정이 가능하다고 볼 수 있으나, 관측지점이 대부분 비슷한 종류의 토양과 토지피복을 가지고 있어 범주형 변수에 따른 변화보다는 NDVI와 같은 연속형 변수에 따른 변화양상을 보였다. CART기법 적용에 따른 부귀의 토양수분의 추정치와 관측치 시계열 자료는 강수사상에 대한 토양수분의 반응과 감쇄거동을 잘 나타내고 있어 모형의 적용성이 높음을 보여준다. 검증 지점으로 선택한 계곡2 지점은 0.7541의 상관성을 보여 생성된 상관식이 적절하다는 것을 보여주었다. 검증 자료를 통하여 검증된 상관식이 용담댐 지역의 토양수분을 추정하는데 적합한 추정이 가능하다고 판단되어 우리나라 전역의 토양수분 분포도를 추정하였다.

### 3.2 우리나라 전역의 토양수분 산정

용담댐 지역에 대해 적용한 산정식을 우리나라 전역을 대상으로 추정하고자 MODIS NDVI자료와 LST, 수자원관리정보시스템(WAMIS)에서 획득한 우리나라 전역에 대한 토지피복, 토양도(유효토심, 종류)자료를 수집하였으며, 우리나라 76개소 기상관측소 중에 자료의 기간이 30년 미만인 관측지점과 지역이 점인 지점들을 제외한 57개 지점의 강수량, 토양온도 자료를 사용하였고 연구기간은 MODIS 위성영상자료와의 비교를 위하여 2003년 1월부터 2008년 12월까지를 선택하였으며, 관측소별 일 강우량을 MODIS 자료 주기와 동일하도록 월 단위로 누적하여 월강우량을 계산하고, 강우량과 같은 지상관측망 자료는 IDW 법을 이용하여 MODIS NDVI 위성영상자료와 동일한 1km 공간해상도를 가진 기상자료로 생성하였다.

CART 기법과 각종 입력 자료를 이용하여 용담댐 지역에 대하여 추정된 토양수분 추정모형을 이용하여 우리나라 2008년 월별 토양수분 분포의 추정 가능성을 검토하였다. Fig. 3은 2008년 한반도 토양수분 추정치의 월별, 공간별 평균을 나타낸 그림으로 월별 토양수분의 공간분포는 용담댐 지역에 적용한 결과와 마찬가지로 식생분포에 따른 지역 특성이 잘 반영된 결과를 보여주었다. 월별 토양수분의 면적평균값의 시계열은 강우 패턴을 전체적으로 잘 반영하고 있는 것으로 판단된다.



(a) Spatial average

(b) Monthly average of soil moisture

Fig. 3. 2008년 우리나라 전역의 평균 토양수분

## 4. 결 론

본 연구에서는 토양수분 변동성에 영향을 주고 공간정보 확보가 용이한 MODIS NDVI, MODIS LST, 토지피복, 토양도(유효토심, 종류), 지상관측망 토양온도, 강우량, 토양수분 등의 보조 자료와 의사결정나무기법

중에 하나인 CART 기법을 활용하여 토양수분자료를 산정하였다.

대상 지역의 3개 지점(부귀, 상전, 천천2) 토양수분자료에 적용한 결과 5개의 최적 산정식을 생성하였으며, 산정식을 이용한 토양수분 추정치와 관측된 토양수분의 상관성을 분석한 결과로 평균 0.747의 상관계수를 보였다. 이는 생성된 토양수분으로 적절한 추정이 가능하다고 볼 수 있으나, 관측지점이 대부분 비슷한 종류의 토양과 토지피복을 가지고 있어 범주형 변수에 따른 변화보다는 NDVI와 같은 연속형 변수에 따른 변화양상을 보였다. 용담댐 유역에 대한 토양수분 추정은 보다 타당한 결과를 보여줌에도 불구하고 적용범위를 우리나라 전역으로 넓혔을 경우에는 전체적인 조건을 만족하기에는 한계가 있다. 이런 문제를 해결하고 CART기법과 가용자료를 이용한 우리나라 전역의 토양수분 추정치의 신뢰성을 높이기 위해서는 보다 많은 토지피복, 식생, 등의 조건에 대한 신뢰성 높은 지상토양수분관측과 원격관측기술의 발달에 따른 항공원격탐사 등에 기초한 기초자료의 확보가 절실하다고 판단된다.

### 감사의 글

본 연구는 국토해양부 첨단도시기술개발사업 - 지능형국토정보기술혁신 사업과제의 연구비지원(07국토정보C03)에 의해 수행되었습니다.

### 참고문헌

1. 김광섭(2006). PSR C-band 및 ESTAR L-band 측정치를 사용한 다중 채널 원격측정 토양수분 자료의 변화도 비교, 대한토목학회 논문집, 대한토목학회, 제26권, 제4B호, pp. 329-334
2. 김광섭, 이을래(2004). 신경망기법과 보조자료를 사용한 원격측정 토양수분자료의 Downscaling기법 개발, 한국수자원학회 논문집, 한국수자원학회, 제37권, 제1호, pp. 21-29
3. 박은주, 황철수, 성정창(2002). 위성영상의 토양수분 정보와 공간적 요인을 고려한 가뭄 민감도 분석, 한국GIS학회지, 한국GIS학회, 제10권, 제3호, pp. 481-492.
4. 유철상(1998) 토양수분 분산도의 시간-공간적 변화에 대한 강우의 영향, 대한토목학회 논문집, 대한토목학회, 제18권, 제II-1호, pp. 23-30.
5. 이가영, 김기훈, 김상현(2004). 설마천 유역의 토양수분 장기 모니터링을 통한 토양수분 시공간 변화양상의 특성화, 2004년 수자원학회 학술발표회 논문집, 한국수자원학회, pp. 209-214.