Journal of the Korean Society of Agricultural Engineers
DOI : https://doi.org/10.5389/KSAE.2019.61.6.123

Vol. 61, No. 6, pp.123~132, November 2019
ISSN 1738-3692 | eISSN 2093-7709

Terra MODIS NDVI 및 LST 자료와 RNN-LSTM을 활용한 토양수분 산정

RNN-LSTM Based Soil Moisture Estimation Using Terra MODIS NDVI and LST

장원진[a] · 이용관[b] · 이지완[c,†] · 김성준[d]
Jang Wonjin · Lee Yonggwan · Lee Jiwan · Kim Seongjoon

ABSTRACT

This study is to estimate the spatial soil moisture using Terra MODIS (Moderate Resolution Imaging Spectroradiometer) satellite data and machine learning technique. Using the 3 years (2015~2017) data of MODIS 16 days composite NDVI (Normalized Difference Vegetation Index) and daily Land Surface Temperature (LST), ground measured precipitation and sunshine hour of KMA (Korea Meteorological Administration), the RDA (Rural Development Administration) 10 cm~30 cm average TDR (Time Domain Reflectometry) measured soil moisture at 78 locations was tested. For daily analysis, the missing values of MODIS LST by clouds were interpolated by conditional merging method using KMA surface temperature observation data, and the 16 days NDVI was linearly interpolated to 1 day interval. By applying the RNN-LSTM (Recurrent Neural Network-Long Short Term Memory) artificial neural network model, 70% of the total period was trained and the rest 30% period was verified. The results showed that the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Nash-Sutcliffe Efficiency were 0.78, 2.76%, and 0.75 respectively. In average, the clay soil moisture was estimated well comparing with the other soil types of silt, loam, and sand. This is because the clay has the intrinsic physical property for having narrow range of soil moisture variation between field capacity and wilting point.

Keywords: RNN-LSTM; MODIS; NDVI; LST; soil moisture; tensorflow

Ⅰ. 서 론

토양수분은 토양 입자 간의 공극에 존재하는 물로 그 양에 따라 토양의 물리적 화학적 속성에 영향을 주는 요소로 지구상 존재하는 물에서 작은 비율을 차지하지만 물순환, 에너지 분포와 지표에서 발생하는 자연현상에 있어 중요한 요소로 작용하며 홍수나 가뭄에도 영향을 끼친다 (Dai et al., 2004). 국내외 수많은 연구자들에 의해 토양수분이 수문순환에 주는 영향에 대한 정량적 해석과 다양한 인자와의 상관성 분석에 관한 연구가 진행되고 있다 (Aubert et al., 2003; Qiu et al., 2003; Zribi et al., 2005). 이러한 상관성 분석 결과에 따르면 토양수분은 강수량, 풍속, 대기온도, 증발산량, 복사량, 습도 등에 영향을 받으며 (Barling et al., 1994; Joo et al., 2010) Normalized Difference Vegetation Index (NDVI)와 Land Surface Temperature (LST)도 토양수분에 높은 상관성을 가지고 있는 것으로 분석되었다 (Farrar et al., 1994; Gillies et al., 1997; Narasimhan et al., 2005).

토양수분 측정은 주로 지상관측과 원격탐사로 이뤄지는데, 지상관측으로는 중량법, 중성자법, Time Domain Reflectometer (TDR)법이 있으며 원격탐사에서는 마이크로파 기반의 센서를 이용하거나 가시/근적외선 위성 영상으로부터 토양수분과의 상관관계를 이용해 산정하는 방법이 있다 (Su et al., 2014). 국내 지상관측 자료의 경우 농업기상정보시스템 (http://weather.rda.go.kr)에서 200개 지점에 대하여 토양수분을 측정 및 제공하고 있으나 제공된 자료는 관측 기간이 짧고 인력, 비용, 관측장비 노후화 등으로 인해 오류 및 결측이 많아 자료의 신뢰성이 부족하고 미계측 지역이 포함된 넓은 공간에서의 활용이 어렵다는 단점이 있다.

인공위성 자료의 경우 넓은 지역의 자료를 연속적으로 수집하지만 공간해상도 (10 km~50 km)와 시간해상도에 한계가 있어 좁은 지역에서 활용하기에는 제약이 있고 지역에 따라 별도의 검보정을 요구하기에 영상처리기법과 통계기법을 적용하여 사용한다. 영상처리기법을 적용한 사례를 살펴

[a] Graduate Student, Department of Civil, Environmental, and Plant Engineering, Konkuk University
[b] Doctoral Student, Department of Civil, Environmental, and Plant Engineering, Konkuk University
[c] PhD Candidate, Department of Civil, Environmental, and Plant Engineering, Konkuk University
[d] Professor, School of Civil, Environmental, and Plant Engineering, Konkuk University
† **Corresponding author**
Tel.: +82-2-444-0186 Fax: +82-2-444-0186
E-mail: closer01@konkuk.ac.kr
Received: October 31, 2019
Revised: November 21, 2019
Accepted: November 22, 2019

보면, 마이크로파 AMSR-E (Advanced Microwave Scanning Radiometer)위성을 이용해 용담댐 유역의 토양수분자료의 상관성분석을 실시한 바 있으며 (Kim and Kim, 2011), 공간해상도를 개선하기 위해 지면변수와 AMSR2 영상에 회귀크리깅을 실시하였고 (Kim et al., 2017a), 지표면 관측자료를 이용한 다운스케일링을 통해 AMSR2 자료를 상세화하였다 (Wang et al., 2009; Kim and Hogue, 2011). 통계 기법을 이용한 토양수분 산정의 경우, Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI, 국내 위성인 천리안의 지면온도와 Multi Linear Regression (MLR)을 이용하여 토양수분을 산정한 바 있으며 (Lee et al., 2017), 위성자료를 달리한 MODIS의 NDVI, LST 자료와 MLR을 이용한 토양별 토양수분 회귀식 개발 (Jung et al., 2017) 및 MODIS 자료를 이용한 토양수분 분포도 작성이 있다 (Hutchinson et al., 2006). 이처럼 국내외에서 인공위성자료를 이용해 토양수분의 산정과 예측을 위해 영상처리기법, 선형회귀분석, 머신러닝, 인공신경망 (Artificial Neural Network, ANN) 등을 적용하는 연구가 활발히 진행 중이다 (Park and Kim, 2011; Kim et al., 2017b).

본 연구에서는 기존 선행연구에서 사용한 0.30~0.76 (R^2). 0.46%~12.21% (RMSE)의 결과를 보인 MLR을 개선하기 위하여 시계열자료 분석에 특화된 순환신경망 (Recurrent Neural Network, RNN)에서 장기 의존성 문제가 보완된 Long Short Term Memory model (LSTM)을 MODIS NDVI, LST 자료와 기상자료에 적용해 토양수분을 산정하고자 한다.

II. 재료 및 방법

본 연구에서는 Python 라이브러리 Tensorflow를 기반으로 한 RNN-LSTM을 이용해 토양수분산정을 알고리즘을 개발하였다. 입력자료는 2015년~2017년의 지상 관측자료 (강수량, 지표온도, 일조시간)를 공간분포해 사용하였으며, 지상관측자료와 동일한 기간의 MODIS 위성의 NDVI, LST를 1 day, 1 km 해상도로 구축하였다. 위성데이터의 공간해상도, 결측값 등에 대한 보정을 위해 조건부합성 (Conditional merging, CM)을 수행하였다 (Fig. 1).

1. 위성데이터 – NDVI, LST

NDVI는 근적외선의 반사율 차이를 이용하는 식생지수로 격자 단위로 식물의 분포 또는 식생활력도를 나타낸다 (Wan et al., 2004). MODIS NDVI product의 경우 좁은 파장 영역의 분광 자료를 활용하기 때문에 대기의 영향이 최소화되어 지표면의 방사해상도가 다른 위성영상보다 좋아 물에 의한 식생의 반응을 효과적으로 반영되는 특징이 있다 (Huete et al., 1999). MODIS NDVI Product는 250 m~1,000 m의 공간해상도, 16일~월 단위의 시간해상도로 합성되어 제공된다. NDVI는 근적외선 영역에서의 반사율 (RNIR)과 가시광선 영역인 적색광의 반사율 (RRED)의 변화로 계산된다 (Eq. 1).

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

NDVI의 값은 -1.0에서 1.0의 범위를 가지며, 물, 눈 구름의 경우에는 수분을 포함하고 있어 근적외선의 반사값이 낮아 0 미만의 음수 값을, 식생이 없는 나지의 경우 근적외와 적색광의 비율이 거의 같아 0에 가까운 값을, 밀도가 높은 식생지역은 근적외의 반사값이 비교적 높아 0.3~0.8의 양의 값을 가진다.

LST는 지표면온도 자료로 MODIS LST의 경우 12번

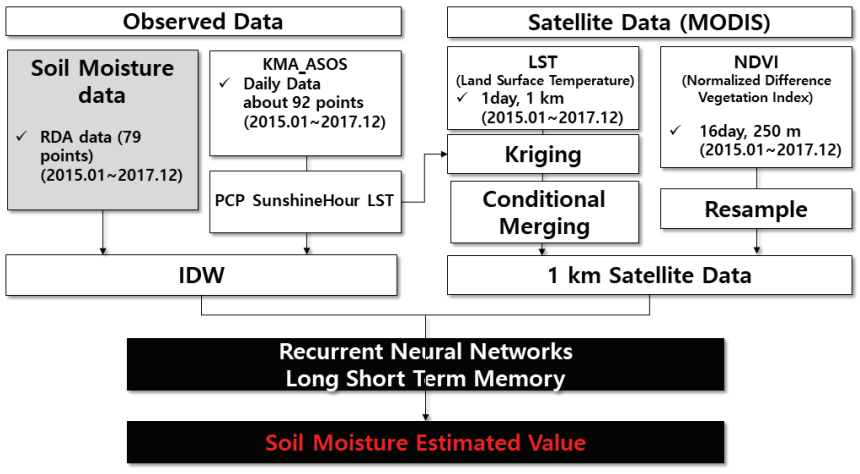


Fig. 1 LSTM soil moisture estimation flow chart

Thermal Infrared (TIR) 열 적외선 대역의 밴드를 이용해 Eq. (2~4), Generalized Split-Window (GSW) 알고리즘, 통계 회귀 방식으로 산정된다 (Wan and Dozier, 1996; Wang, 1999).

$$T_s = C + (A_1 + A_2\frac{1-\epsilon}{\epsilon} + A_3\frac{\Delta\epsilon}{\epsilon^2}) \times \frac{T_{31}+T_{32}}{2} + (B_1 + B_2\frac{1-\epsilon}{\epsilon} + B_3\frac{\Delta\epsilon}{\epsilon^2}) \times \frac{T_{31}-T_{32}}{2} \quad (2)$$

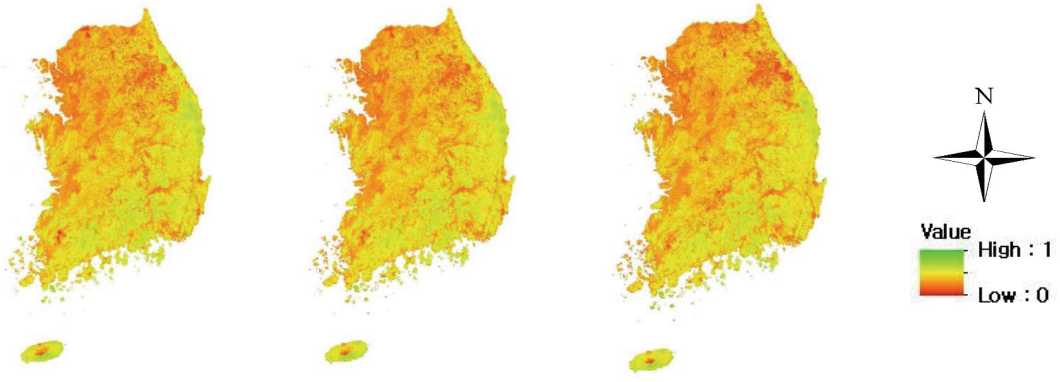
$$\epsilon = (\epsilon_{31} + \epsilon_{32})/2 \quad (3)$$

$$\Delta\epsilon = \epsilon_{31} - \epsilon_{32} \quad (4)$$

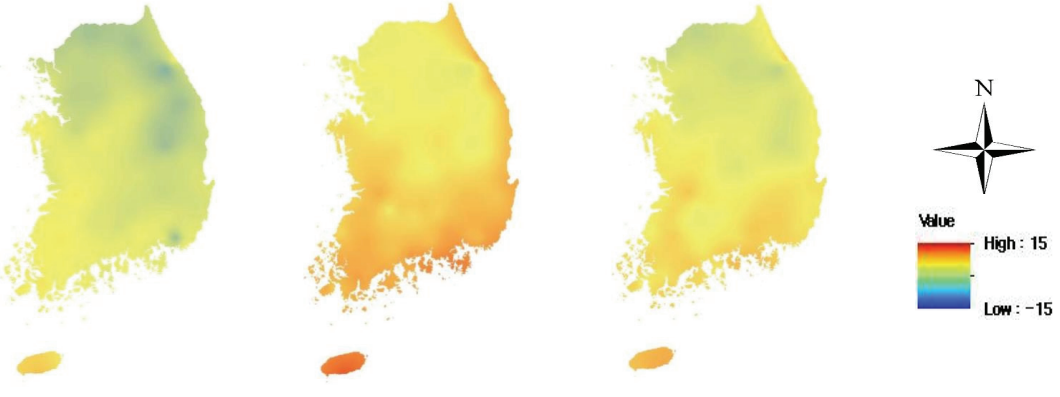
여기서 T_{31}, T_{32}는 MODIS 31번 32번 밴드의 밝기 온도, ϵ_{31}, ϵ_{32}는 MODIS 31번 32번 밴드의 표면 방사율, 그리고 C, A_1, A_2, A_3, B_1, B_2, 그리고 B_3는 회귀 계수이다. LST는 순 복사 에너지양에 비례하기 때문에 지표의 수분 상태를 반영하는 인자로 활용이 되며 이와 같은 이유로 토양수분의 상태를 파악하고 지표 피복에 따른 가뭄을 탐지하는 등과 같이 널리 활용되고 있다 (Park, 2003; Wan et al., 2004).

NDVI와 LST자료를 활용하기 위해 Land Processes Distributed Active Archive Center (LPDAAC, https://lpdaac.usgs.gov/)에서 제공하는 MOD13Q1 16일 간격의 250 m 공간해상도의 NDVI를 구축 후 resample를 이용해 해상도를 1,000 m로 통일하였으며, LST자료와 시간해상도를 일치시키기 위해 선형보간을 통해 일자료를 생성하였다 (Fig. 2(a)). LST의 경우 MOD11A1 1일 간격의 1,000 m 공간해상도 자료를 기상청 지상관측소 지표온도와 조건부합성 (CM)을 적용해 구름의 간섭에 의한 데이터의 오류와 결측을 보정하였다 (Fig. 2(b)).

조건부합성은 기상분야에서 강우량 산정을 위해 사용되었던 기법으로 레이더나 위성자료에서 측정된 분포패턴과 직접 측정된 강우량을 합성하는데 사용된다. 직접 측정된 자료는 지질 통계학적 접근 방법인 크리깅 (Kriging)기법으로 공간분



(a) Linear interpolated NDVI



(b) Conditional merged LST (℃)

Fig. 2 Terra MODIS data

포 후, 해당 자료와 원격탐사자료와의 편차를 이용해 보다 현실적인 공간자료를 생산한다 (Kim and Jung, 2016). 본 연구에서는 92개 기상 관측 지점에서 수집된 지표면 온도자료를 공간분포하고 MODIS LST자료로부터 지상관측소 지점과 일치하는 지점에서의 LST를 추출하고 추출된 지점 값으로부터 크리깅 기법을 이용해 공간보간을 실시하였다. 그 뒤 잔차를 산정하고, 산정된 잔차에서 지상 LST 자료를 합성해 최종적으로 조건부 결합된 공간 LST자료를 산정하였다.

2. 토양수분자료

토양수분자료는 농촌진흥청 (weather.rda.go.kr)에서 제공되는 104개 지점의 농업기상자료를 활용하여 토양수분자료를 구축하였다. 이 중 결측치나 데이터의 오차가 존재하는 지점을 제외하고 78개소의 자료를 사용하였다 (Table 1). 토양수분 자료는 용적수분함량 (Volumetric Water Content) 또는 TDR방식으로 측정되었으며 지중 10 cm와 30 cm 깊이의 수분자료로 매 시간마다 두 자료의 산술평균으로 토양수분이 산정된다. Fig. 3은 최종 선정된 78개소의 위치정보와 대표 6개소 (OM: 오미동, PB: 포항시 북구, HJ: 황지동, AH: 압해면, SY: 소이면, CI: 처인구)에 대해 수집된 토양수분자료를 도시한 것이다.

3. RNN – LSTM

기계학습 (Machine learning)은 인공 지능의 한 분야로, 컴퓨터가 데이터를 학습할 수 있도록 하는 알고리즘과 관련 기술들을 말한다. 하드웨어의 발전에 따라 모든 분야에서의 데이터 활용 가능성을 높여주고 있다. 대표적인 학습알고리즘으로는 ANN, DNN (Deep Neural Network), RNN, CNN (Convolutional Neural Network), RBM (Restricted Boltzmann Machine) 등이 있다.

여기서 RNN은 셀에서 입력값을 받아 학습하고 셀의 출력값이 다음 스텝에서 다시 입력값이 되는 순환구조를 가지는 인공신경망의 한 종류로 은닉계층에 과거의 데이터를 기억해 학습을 진행해 자연어나 음성신호, 주식, 기상과 같이 연속성을 가지는 시계열 데이터에서 우수한 성능을 보이는 모델이다 (Fig. 4(a)). 특히 시계열 분석에 있어 빠르고 정확한 학습의

Table 1 Soil moisture observation points location information

Location	Latitude	Longitude	A.D.[a]	Location	Latitude	Longitude	A.D.	Location	Latitude	Longitude	A.D.
BGS	35.21	128.96	BS	PB	36.11	129.31	GB	HD	37.71	127.86	GW
NI	36.59	127.51	CB	GP	37.85	127.5	GG	HC2	38.11	127.7	GW
BY	37.16	128.18	CB	GJ2	37.82	126.97	GG	HJ	37.16	128.98	GW
SY	36.91	127.76	CB	DW	37.25	126.59	GG	HR	38.18	128.08	GW
SC	36.21	127.72	CB	BD	37.22	126.95	GG	HU	37.3	127.92	GW
OC	36.72	127.47	CB	SR	37.3	126.84	GG	MJ	36.01	127.68	JB
OC2	36.3	127.6	CB	YP	37.51	127.51	GG	BR	35.75	126.85	JB
JC	36.85	127.43	CB	JR	37.27	127.43	GG	YB	36.43	127.44	JB
HD2	36.62	127.42	CB	CI	37.17	127.3	GG	JS2	35.62	127.51	JB
GA	36.26	126.84	CN	GJ	34.86	128.58	GN	JW	35.62	126.9	JB
DM	36.26	127.28	CN	GC	35.67	127.92	GN	JA	35.76	127.44	JB
SA	36.74	126.81	CN	GS	34.99	128.33	GN	GR	35.2	127.46	JN
YC	36.83	126.97	CN	GD	34.9	128.41	GN	GN	34.51	126.3	JN
CY	36.43	126.8	CN	DJ	35.55	128.48	GN	GS2	35.28	126.47	JN
TA	36.73	126.31	CN	SN	35.45	128.76	GN	DJ	34.83	126.67	JN
HS	36.6	126.58	CN	YH	35.04	128.07	GN	AH	34.9	126.34	JN
DG	35.91	128.25	GB	JR	35.08	127.77	GN	OG	35.27	127.3	JN
MH	36.92	129.38	GB	CJ	35.21	128.12	GN	OC2	34.58	126.65	JN
BG	36.57	129.4	GB	GG	37.53	127.96	GW	JS	35.32	126.8	JN
SP	37.05	129	GB	GNM	37.34	129.01	GW	JRDA	35.02	126.83	JN
SS	36.23	128.29	GB	GD2	37.38	129.23	GW	JS3	34.79	127.65	JN
AJ	36.85	128.56	GB	DP	38.17	128.6	GW	HG	35.05	126.54	JN
YM	36.04	128.38	GB	BP	37.49	129.12	GW	HC	34.97	127.07	JN
YC2	36.68	128.46	GB	YM2	37.38	128.4	GW	HG	35.03	126.45	JN
OM	35.99	128.93	GB	YY	37.17	128.48	GW				
JI	35.82	128.81	GB	IJ	38.06	128.17	GW				
CS	36.45	128.17	GB	JB	37.67	128.59	GW				

[a] Administrative district: BS (Busan), CB (Chungcheongbuk-do), CN (Chungcheongnam-do), GB (Gyungsanbuk-do), GG (Gyeonggi-do), GN (Gyungsannam-do), GW (Gangwon-do), JB (Jeollabuk-do), JN (Jeollanam-do).

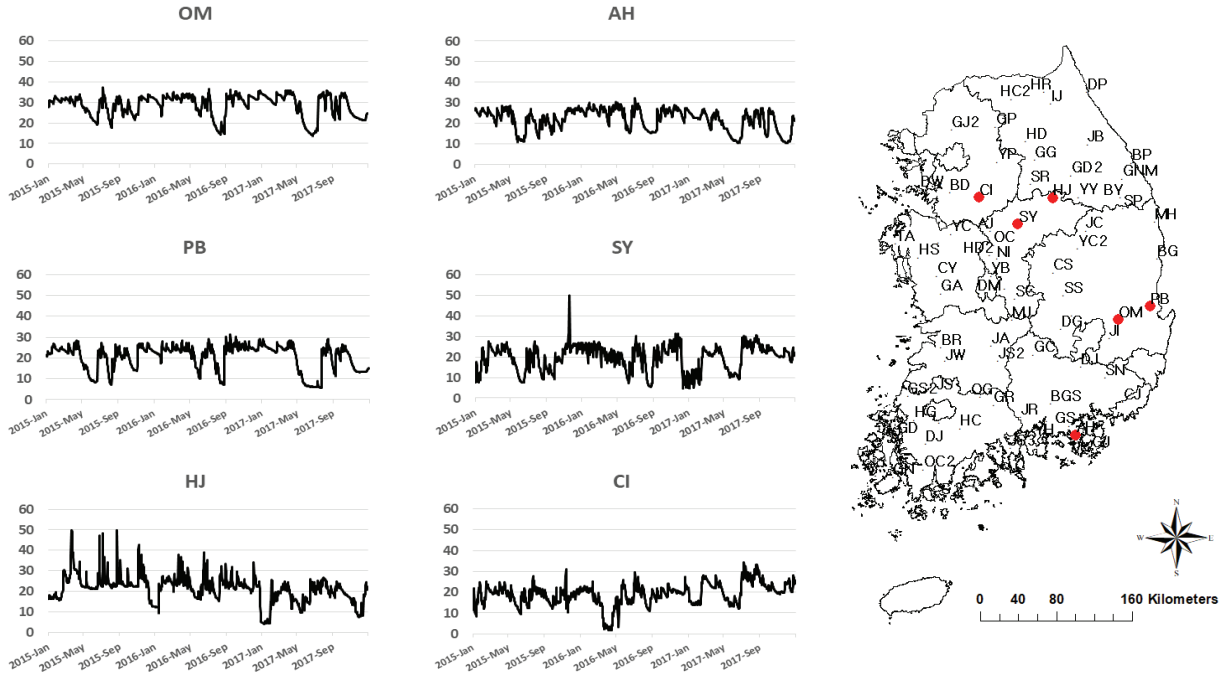
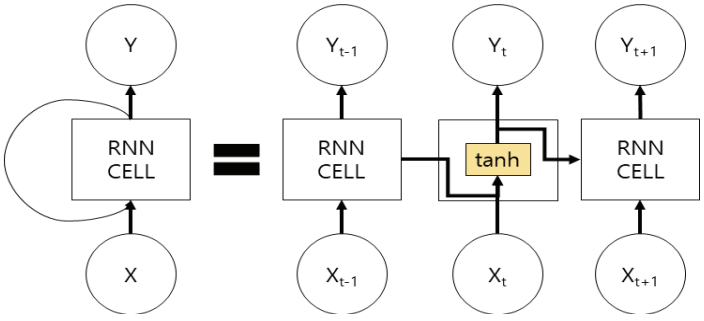
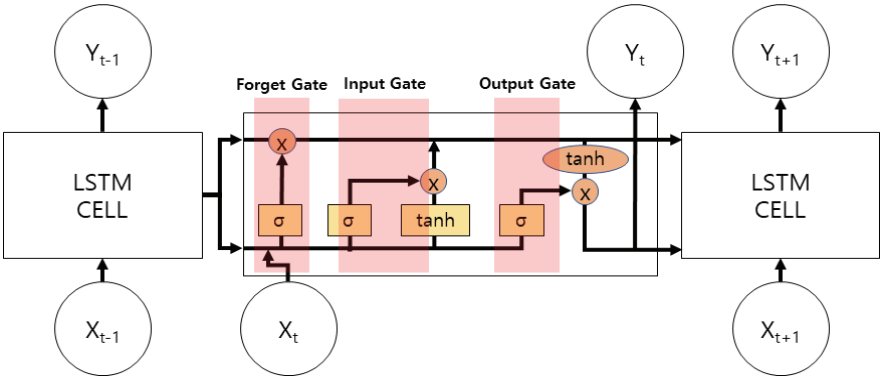


Fig. 3 10~30 cm Soil moisture monitoring station (6 points)



(a) RNN structure



(b) LSTM structure

Fig. 4 Machine learning structure (Gers et al., 2002)

특징을 가진다. 하지만 이러한 장점에도 불구하고 입력자료가 많거나 학습횟수를 과도하게 설정하면 기울기 소실 (Vanishing gradient)에 따른 학습의 저하 및 불가, 또는 속도 저하가 발생하게 된다. 이러한 장기 의존성에 대한 문제를 해결하기 위해 RNN에 LSTM 구조를 도입하였다.

LSTM은 기존 RNN Cell에 입력 게이트 (input gate), 망각 게이트 (forget gate), 출력 게이트 (output gate)가 추가된 구조이다 (Fig. 4(b)). 여기서 입력 게이트는 sigmoid 함수를 이용해 이전 셀에서 들어온 입력 값을 받아드리는 정도를, 망각 게이트는 sigmoid와 tanh 함수를 이용해 이전 셀의 학습결과를 얼마나 가져올지, 출력 게이트는 셀을 업데이트하고 마지막 sigmoid 함수와 셀 스테이트 출력결과를 tanh 함수 이용해 얼마나 출력할지를 결정한다 (Shin et al., 2017). 이러한 구조를 통해 반복된 학습에서 발생하는 기울기 소멸문제를 해결하고 장기 의존성을 유지할 수 있다 (Gers et al., 2002). 본 연구에서는 LSTM을 사용하기 위해 기계학습과 딥러닝을 위해 구글에서 배포하는 Python TensorFlow를 사용하였다.

III. 결과 및 고찰

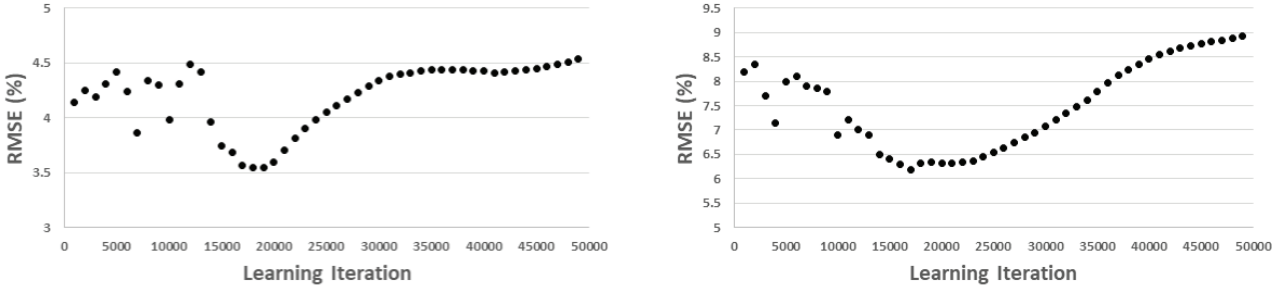
1. LSTM 모형 구축

토양수분 산정을 위해 입력변수 (NDVI, LST, 일조시간, 강수량, 토양수분) 5개, 출력변수 (다음날 토양수분) 1개를 사용하였으며 여기서, 입력변수는 Lee et al. (2017)의 선행연구결과를 참고하였다. Lee et al. (2017)은 Stepwise 방법을 이용하여 NDVI, LST, 일조시간, 5일 강수량을 순차 적용하여 회귀식에 가장 지배적인 영향을 주는 인자를 도출한 결과 모든 독립변수를 사용하였을 때 가장 좋은 상관성을 나타냄을 확인한 바 있다. 산정 알고리즘은 3개의 층을 가지는 다층구조 LSTM을 Python 오픈소스 라이브러리 Tensorflow를 이용하여 작성되었으며 은닉층 (Hidden Layer)의 개수는 10개, 배치크기 (Sequence length)는 5, 학습 정도의 지표는 평균제곱근오차 (Root Mean Square Error, RMSE)를 최적화 방법은 지표의 감소속도가 빠른 AdamOptimizer를 사용하였다. 여기서 배치크기는 토양수분 산정을 위해 입력되는 시계열 자료의 개수로 값이 3인 경우 3일간의 자료를 이용해 하루의 토양수분을 산정을 의미한다. 본 연구에서의 배치 크기는 선행연구결과를 토대로 선행토양함수조건을 고려하기 위해 5일로 설정하였다 (Jung et al., 2017).

기계학습의 적용 시 가장 중요한 인자는 학습횟수이다. 학습횟수가 과도하게 클 경우, 과적합 (Overfitting) 문제가 발생하게 된다. 본 연구에서는 과적합 문제를 방지하기 위해 전국에 적용하기 앞서 관측자료의 품질이 가장 좋은 3개의 포인트 (Fig. 5(a))와 나쁜 3개의 포인트 (Fig. 5(b))에 대하여 학습횟수에 따른 RMSE 변화를 분석하였다. 1,000회부터 50,000회 학습을 실시하였고 앞선 알고리즘의 조건에서는 학습횟수가 20,000회 이상 넘어가게 되면 학습자료에 과적합이 나타나 검증결과의 RMSE가 증가하는 것을 확인하였다. 따라서, 본 연구에서는 전체 78개 지점에 대하여, 동일기간 (2015.01.01.~2017.02.03.)에 20,000번 학습시켰다.

2. LSTM 모형 적용 결과

본 연구에서는 농촌진흥청의 78개 지점에 대하여 LSTM 알고리즘을 적용하여 토양수분자료를 산정하였다. 토양수분 산정에 사용한 입력자료는 조건부합성을 통해 결측값이 보간된 LST, 선형보간으로 생성된 일별 NDVI, 일조시간, 강수자료를 사용하였다. LSTM 모형의 적용성 평가를 위해 결정계수 (coefficient of determination, R^2), RMSE, NSE 모형효율성계수 (Nash-Sutcliffe Efficiency, NSE)로 비교하였다. 여기서 R^2, NSE의 경우는 1에 가까울수록, RMSE의 경우 0에 가까울수록 실측값과 산정값이 좋은 상관성을 가짐을 의미한다. Table 2는 학습데이터를 이용해 학습된 LSTM 알고리즘을 각



(a) (b)

Fig. 5 RMSE change according to learning iteration

Table 2 Validation results for estimated soil moisture

LOC	R^2	NSE	RMSE	Type	LOC	R^2	NSE	RMSE	Type	LOC	R^2	NSE	RMSE	Type
GP	0.83	0.82	1.77	Clay	SN	0.86	0.84	1.28	Silt	JW	0.86	0.83	2.09	Sand
GJ	0.59	0.57	6.21	Sand	SR	0.82	0.81	2.34	Silt	JS3	0.82	0.82	2.38	Loam
GC	0.70	0.65	3.17	Silt	SP	0.78	0.77	2.63	Sand	JR	0.70	0.67	2.92	Silt
GS	0.71	0.68	4.11	Silt	SS	0.69	0.64	2.73	Loam	JB	0.89	0.88	2.13	Loam
GG	0.80	0.79	3.32	Silt	SY	0.53	0.53	4.33	Sand	JA	0.70	0.66	4.37	Silt
GD	0.68	0.63	5.37	Loam	SA	0.78	0.74	2.54	Silt	CJ	0.77	0.77	2.56	Sand
GJ2	0.64	0.64	3.70	Loam	SC	0.75	0.74	2.42	Loam	JC	0.80	0.71	2.32	Silt
GR	0.83	0.82	2.15	Silt	AJ	0.81	0.80	2.38	Silt	CI	0.58	0.47	3.29	Silt
GN	0.77	0.75	3.04	Loam	AH	0.95	0.94	1.23	Sand	CY	0.77	0.73	2.85	Loam
GS2	0.84	0.84	2.33	Clay	YM	0.89	0.88	2.38	Loam	CS	0.88	0.87	1.93	Sand
GNM	0.73	0.69	2.35	Sand	YP	0.70	0.68	2.86	Silt	TA	0.87	0.84	3.28	Silt
GA	0.69	0.68	4.37	Loam	YM2	0.87	0.85	2.70	Silt	PB	0.93	0.93	1.83	Loam
GD2	0.88	0.87	1.47	Loam	YC	0.75	0.71	2.60	Silt	HG	0.91	0.90	1.87	Clay
NI	0.86	0.84	2.45	Loam	YY	0.61	0.45	3.91	Silt	HC	0.86	0.85	2.68	Loam
DW	0.64	0.61	3.58	Silt	YC2	0.83	0.80	2.72	Silt	HG	0.80	0.79	1.67	Silt
DG	0.80	0.69	2.89	Silt	OG	0.80	0.79	2.45	Loam	HS	0.76	0.73	3.14	Clay
DJ	0.85	0.83	1.72	Clay	OM	0.97	0.96	1.21	Loam	HD	0.80	0.78	2.41	Silt
DP	0.73	0.70	2.02	Loam	OC	0.86	0.84	2.17	Loam	HC2	0.84	0.82	2.09	Silt
DJ	0.82	0.82	2.60	Clay	OC2	0.78	0.75	3.88	Loam	HJ	0.38	0.34	5.85	Loam
DM	0.70	0.64	2.50	Silt	OC2	0.83	0.81	2.49	Silt	HR	0.69	0.68	4.87	Silt
MH	0.69	0.68	2.78	Silt	YH	0.80	0.79	2.28	Sand	HD2	0.89	0.85	2.08	Silt
MJ	0.87	0.85	2.54	Silt	YB	0.74	0.70	2.67	Silt	HU	0.63	0.61	2.68	Sand
BG	0.76	0.73	3.08	Sand	IJ	0.80	0.75	2.58	Loam	Clay	0.85	0.79	2.32	–
BD	0.89	0.87	1.92	Clay	JI	0.91	0.88	2.28	Loam	Loam	0.78	0.77	2.77	
BY	0.67	0.18	4.91	Clay	JS	0.90	0.89	1.76	Sand	Silt	0.76	0.71	2.80	
BR	0.71	0.66	2.58	Sand	JS2	0.81	0.80	3.08	Silt	Sand	0.77	0.75	2.68	
BGS	0.85	0.85	1.61	Loam	JR	0.64	0.64	3.72	Loam	AVG	0.78	0.75	2.76	–
BP	0.85	0.84	2.11	Silt	JRDA	0.89	0.89	1.79	Sand					

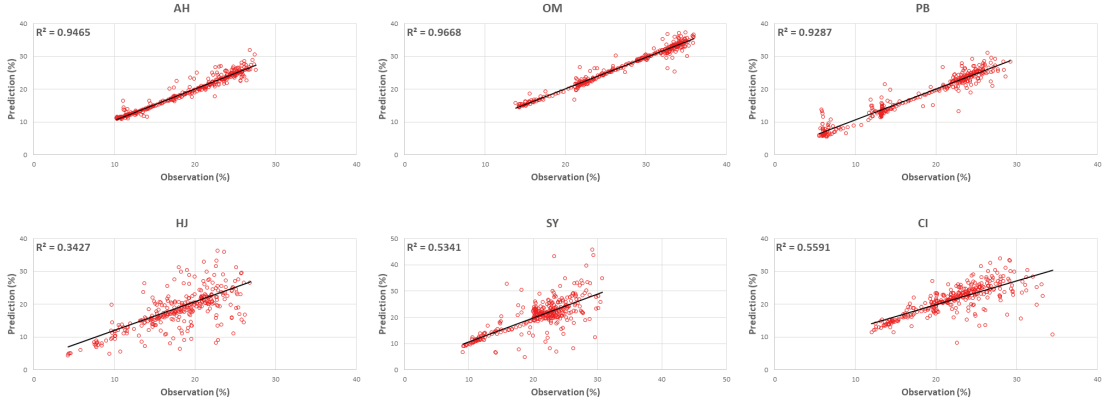


Fig. 6 Soil moisture Validation scatter plot (Top 3 points, Low 3 points)

관측소의 검증데이터에 적용하여 산정된 토양수분 값의 통계결과로 76개 지점에 대한 평균 R^2는 0.78, NSE와 RMSE는 각각 0.75, 2.76%를 보여 모든 지점에서 통계적으로 유의한 수준인 것으로 분석되었다.

Fig. 6과 Fig. 7은 학습결과가 가장 좋은 상위 3개 지점과 하위 3개의 지점에 대해 관측자료와의 1:1 및 시계열 그래프를 도시한 것이다. 학습결과가 가장 우수한 것으로 나타난 오미동 (OM), 압해읍 (AH), 포항 (PB)의 경우 R^2, NSE, RMSE가 각각 0.93~0.97, 0.93~0.96, 1.83~1.21%의 범위를 보이며 관측과 산정값이 선형관계를 보였으며 토양수분의 시계열 변화

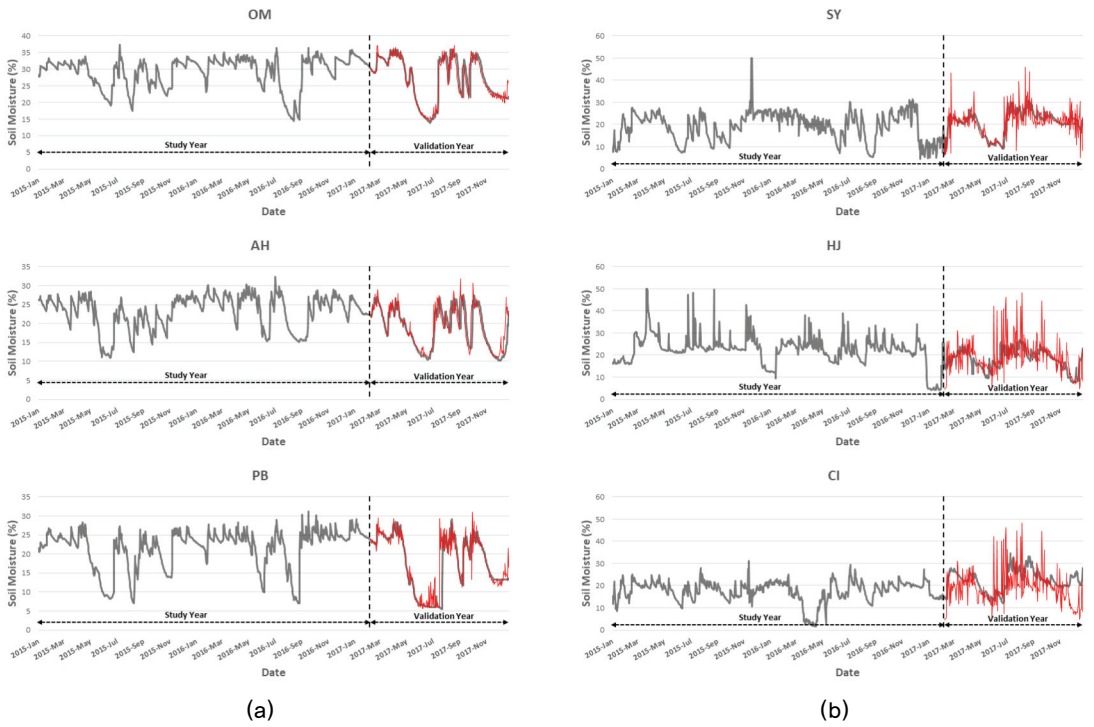


Fig. 7 Soil moisture estimation results graph

또한 관측자료와의 패턴을 매우 잘 모의하는 것으로 분석되었다. 반면 하위 3개 지점인 소이면 (SY), 황지동 (HJ), 처인구 (CI)의 경우 각각 R^2, NSE, RMSE가 각각 0.38~0.58, 0.34~0.53, 3.29~5.85% 범위를 보이며 선형관계를 보이기는 하나 값이 과대, 과소 추정되는 경우가 많고 패턴 앞선 3개 지점보다 완벽히 모의 되지 못함을 확인하였다 (Fig. 6, Fig. 7(b)).

결과가 좋은 상위지점과 하위지점의 학습결과의 차이의 원인을 살펴보면, 상위 지점의 경우 (Fig. 7(a)) 학습년도와 검증년도의 패턴이 유사하지만 하위 지점 특히, 황지동 (HJ) 지점의 경우 학습년도에서는 토양수분 실측값이 주로 20~30%의 범위에서 분포하고 있지만 검증년도에서는 10~25%의 범위로 토양수분이 하강하는 패턴의 변화가 발생해 학습된 알고리즘으로 산정된 값이 검증년도의 패턴을 따라가지 못하는 것으로 보인다. 또한, 앞선 상위지점보다 하위지지점은 이상치의 수가 많아 학습에 영향을 주어 검증 값에도 과도한 이상치가 발생 된 것으로 판단된다 (Fig. 7(b)).

또한, 토양수분은 토양 종류 혹은 속성에 따라 그 범위와 변화 특성이 다른데 이에 대한 차이를 분석하기 위해 검증결과를 토양별로 분석한 결과 Clay가 가장 우수하게 나타났으며, 배수능이 떨어질수록 즉, 토양이 토양수분의 변화에 비교적 덜 민감한 경우 알고리즘의 정확도가 높게 나타났다 (Table 2).

IV. 결 론

본 연구에서는 윈도우 기반의 파이썬 (Python) 3.7버전 Tensorflow를 이용하여 LSTM을 구축하였다. LSTM의 입력자료로 NDVI, LST, 강수량, 일조시간과 토양수분 총 5개의 자료를 활용하였으며 여기서 위성자료의 경우 LST는 조건부 합성을 통해 결측값을 보간하였으며, NDVI의 경우 16일 시간해상도를 선형보간을 통해 일일 NDVI자료로 변환하였고 토양수분과 기상 자료의 경우 농촌진흥청에서 제공하는 104개 지점 중 78개의 10 cm~30 cm 깊이의 실측자료와 기상청에서 제공하는 강수량, 일조시간을 사용하여 공간분포하였다. 5일간의 입력자료를 이용해 다음날 토양수분을 산정하도록 하였고, 각 토양수분 관측소마다 2015~2017년자료의 총 기간 중 70%에 해당하는 기간을 20000번씩 반복 학습시켰으며 30%를 검증기간으로 활용하였다. 학습된 LSTM 알고리즘을 토양수분 관측소별로 적용한 결과 평균 결정계수 (R^2) 0.78, 모형효율계수 (NSE) 0.75, 평균제곱근편차 (RMSE) 2.76%를 보여 모든 지점에서 높은 정확도를 보였고, 특히 토양의 수분 변화에 비교적 덜 민감한 토양인 Clay에 가까울수록 상관성이 높게 나타났다.

하지만, 관측값에 많은 수의 이상치와, 학습년도와 검증년도의 토양수분 패턴의 변화가 크게 발생했던 관측소의 경우 알고리즘 학습에 영향을 주어 토양수분 산정결과가 검증 값

과 상관성이 떨어지는 것으로 관찰되었다. 이러한 학습데이터의 신뢰도와 양이 부족했음에도 불구하고 토양수분의 변화를 잘 모의하고 있으나 추후 연구에서는 알고리즘 정확도 개선과 미계측 지역의 토양수분 산정을 위해 다음과 같은 수정이 필요할 것으로 보인다. 첫째, 앞선 황지동, 처인구 관측소와 같은 패턴의 변화에 따른 알고리즘 성능 저하를 방지하기 위해 학습자료를 기존 방식인 시간 순서에 따른 분할이 아닌 학습기간에 보정기간의 일부분을 혼합하거나 무작위추출 등을 이용해 학습자료를 생성하거나 토양 유형에 따른 토양수분 함양량을 이용해 추가적인 품질관리를 실시하여야 한다. 둘째, 현재 연구에서 사용된 입력자료는 LST, NDVI, 일조시간, 강수량으로 선행연구에서 제시된 4개의 자료만을 사용하였지만, 토양수분은 풍속, 토양의 물리적 특성 등 다양한 요소에 의해 영향을 받으므로 알고리즘의 성능을 개선과 불확실성을 제거하기 위해 입력자료를 추가해야 한다. 셋째, 알고리즘을 통한 미계측 지역의 토양수분을 산정하기 위해서는 토양수분 관측소들을 토양에 따라 세분류하고 토양별 학습 후 미계측 유역에 해당하는 토양에 해당하는 결과를 적용해야 높은 정확도의 토양수분 결과를 얻을 수 있을 것이다. 본 연구에서 개발된 알고리즘은 지속적인 학습과 검증을 통해 정확도 향상이 가능하며 해당 기술을 바탕으로 향후 기후변화 시나리오를 적용한 토양수분 변화 예측, 미계측유역의 토양수분 산정이 가능할 것으로 판단된다.

감사의 글

본 연구는 환경부의 물관리연구사업 (79617)에서 지원받았습니다.

REFERENCES

1. Aubert, D., C. Loumagne, and L. Oudin, 2003. Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model. *Journal of Hydrology* 280(1-4): 145-161. doi:10.1016/S0022-1694(03)00229-4.
2. Barling, R. D., I. D. Moore, and R. B. Grayson, 1994. A quasi-dynamic wetness index for characterizing the spatial distribution of zones of surface saturation and soil water content. *Water Resources Research* 30(4): 1029-1044. doi:10.1029/93WR03346.
3. Dai, A., K. E. Trenberth, and T. Qian, 2004. A global dataset of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming. *Journal of Hydrometeorology* 5(6): 1117-1130. doi:10.1175/JHM-386.1.
4. Farrar, T. J., S. E. Nicholson, and A. R. Lare, 1994. The influence of soil type on the relationships between NDVI, rainfall, and soil moisture in semiarid Botswana. II. NDVI response to soil moisture. *Remote Sensing of Environment* 50(2): 121-133. doi:10.1016/0034-4257(94)90039-6.
5. Gers, F. A., N. N. Schraudolph, and J. Schmidhuber, 2002. Learning precise timing with LSTM recurrent networks. *Journal of Machine Learning Research* 3(Aug): 115-143.
6. Gillies, R. R., W. P. Kustas, and K. S. Humes, 1997. A verification of the 'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the Normalized Difference Vegetation Index (NDVI) and surface. *International Journal of Remote Sensing* 18(15): 3145-3166. doi:10.1080/014311697217026.
7. Huete, A. R., C. Justice, and W. J. D. van Leeuwen, 1999. *MODIS Vegetation Index (MOD13) Algorithm Theoretical Basis Document Version 3*.
8. Hutchinson, J. S., T. J. Vought, and S. L. Hutchinson, 2006. Continuous soil moisture mapping using MODIS NDVI and LST products. *Papers and Proceedings of Applied Geography Conferences* 29: 140.
9. Joo, J. Y., M. H. Choi, S. W. Jung, and S. O. Lee, 2010. Prediction of soil moisture using hydrometeorological data in selmacheon. *Journal of the Korean Society of Civil Engineers* 30(5B): 437-444 (in Korean).
10. Jung, C. G., Y. G. Lee, Y. H. Cho, and S. J. Kim, 2017. A study of spatial soil moisture estimation using a multiple linear regression model and MODIS Land surface temperature data corrected by conditional merging. *Remote Sensing* 9(8): 870. doi:10.3390/rs9080870.
11. Kim, B. S., and B. K. Jung, 2016. Flood simulation using the gauge-adjusted radar rainfall and physics-based distributed hydrologic model. *Hydrological Processes* 22(22): 4400-4414. doi:10.1002/hyp.7043.
12. Kim, D. S., N. W. Park, N. Kim, K. J. Kim, S. J. Lee, Y. H. Kim, J. W. Kim, D. Y. Shin, Y. H. Cho, and Y. W. Lee, 2017a. Downscaling advanced microwave scanning radiometer 2 (AMSR2) soil moisture data using regression-kriging. *Journal of the Korean Cartographic Association* 17(2): 99-110 (in Korean). doi:10.4172/2169-0049.1000139.
13. Kim, G. S., and J. P. Kim, 2011. Correlation analysis between soil moisture retrieved from satellite images and

ground network measurements. *Journal of the Korean Association of Geographic Information Studies* 14(2): 69-81 (in Korean). doi:10.11108/kagis.2011.14.2.069.

14. Kim, J. Y., and T. S. Hogue, 2011. Improving spatial soil moisture representation through integration of AMSR-E and MODIS products. *IEEE Transactions on Geoscience and Remote Sensing* 50(2): 446-460. doi:10.1109/TGRS.2011.2161318.
15. Kim, Y. H., G. J. Kim, S. J. LEE, J. W. Kim, and Y. W. Lee, 2017b. Deep learning-based retrieval of daily 500-m soil moisture for south korea. *Journal of The Korean Cartographic Association* 17(3): 109-121 (in Korean). doi:10.16879/jkca.2017.17.3.109.
16. Lee, Y. G., C. G. Jung, Y. H. Jo, and S. J. Kim, 2017. Estimation of soil moisture using multiple linear regression model and COMS land surface temperature data. *Journal of the Korean Society of Agricultural Engineers* 59(1): 11-20(in Korean). doi:10.5389/KSAE.2017.59.1.011.
17. Narasimhan, B., R. Srinivasan, J. G. Arnold, and M. Di Luzio, 2005. Estimation of long-term soil moisture using a distributed parameter hydrologic model and verification using remotely sensed data. *Transactions of the ASAE* 48(3): 1101-1113. doi:10.13031/2013.18520.
18. Park, J. A., and G. S. Kim, 2011. Estimation of spatial distribution of soil moisture at Yongdam Dam watershed using artificial neural networks. *Journal of the Korean Geographical Society* 46(3): 319-330 (in Korean).
19. Park S. Y., 2003. Evaluation of MODIS land surface temperature as an indicator of the climatic water budget in the central great plains. *The Geographical Journal of Korea* 37(3): 257-271 (in Korean).
20. Qiu, Y., B. Fu, J. Wang, and L. Chen, 2003. Spatiotemporal prediction of soil moisture content using multiple-linear regression in a small catchment of the Loess Plateau, China. *CATENA* 54(1-2): 173-195. doi:10.1016/S0341-8162(03)00064-X.
21. Shin, D. H., K. H. Choi, and C. B. Kim, 2017. Deep learning model for prediction rate improvement of stock price using RNN and LSTM. *Journal of Korean Institute of Information Technology* 15(10): 9-16 (in Korean). doi: 10.14801/jkiit.2017.15.10.9.
22. Su, S. L., D. N. Singh, and M. S. Baghini, 2014. A critical review of soil moisture measurement. *Measurement* 54: 92-105. doi:10.1016/j.measurement.2014.04.007.
23. Wang, L., J. Wen, T. Zhang, Y. Zhao, H. Tian, X. Wang, R. Liu, J. Zhang, and S. Lu, 2009. Surface soil moisture estimates from AMSR-E observations over an arid area. *Hydrology and Earth System Sciences Discussions* 6(1): 1055-1087.
24. Wang, Z. M., 1999. MODIS land-surface temperature algorithm theoretical basis document (LST ATBD). *Institute for Computational Earth System Science, Santa Barbara,* 75.
25. Wan, Z., and J. Dozier, 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing* 34(4): 892-905. doi:10.1109/36.508406.
26. Wan, Z., P. Wang, and X. Li, 2004. Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains. *International Journal of Remote Sensing* 25(1): 61-72. doi:10.1080/0143116031000115328.
27. Zribi, M., N. Baghdadi, N. Holah, and O. Fafin, 2005. New nethodology for soil surface moisture estimation and its application to ENVISAT-ASAR multi-incidence data inversion. *Remote Sensing of Environment* 96(3-4): 485-496. doi:10.1016/j.rse.2005.04.005.