



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

인공지능과 기상 환경인자를 이용한 깊이별  
밭 토양온도 예측 모델 개발 및 정확성 평가

Development and accuracy evaluation of field  
soil temperature prediction model by depth  
using artificial intelligence and  
meteorological parameters

2022년 2월

서울대학교 대학원  
바이오시스템공학과  
김 대 현

인공지능과 기상 환경 인자를  
이용한 깊이별 밭 토양온도 예측  
모델 개발 및 정확성 평가

Development and accuracy evaluation of  
field soil temperature prediction model by  
depth using artificial intelligence and  
meteorological parameters

지도교수 이 중 용

이 논문을 공학석사 학위논문으로 제출함  
2021년 12월

서울대학교 대학원  
바이오시스템공학과  
김 대 현

김대현의 석사 학위논문을 인준함  
2022년 1월

위 원 장	<u>박 영 준</u>	(인)
부위원장	<u>이 중 용</u>	(인)
위 원	<u>김 준 용</u>	(인)

## 국문 초록

각종 첨단 기술을 활용하여 기상, 토양 환경 및 작물 정보를 분석하고 자원의 사용을 최소화하며 생산성을 극대화하는 정밀농업 기술이 노지에서도 적용되어야 할 필요가 있지만 외부 기상의 영향을 많이 받아 환경제어가 어려운 노지의 경우 발생하는 장비의 비용에 비해 효율성이 낮아 실제 농업환경에 적용되는 사례가 적다.

이 중 토양은 농작물 성장에 직접적으로 관여하므로 작물의 정상적 성장을 위해 토양 환경 데이터를 습득하는 것은 필수적이며, 토양온도 데이터의 경우 비료 시비, 관수, 농작업 일정 등 농업활동의 여러 의사결정을 하는데 필요한 중요한 정보이기 때문에 깊이에 따른 데이터를 습득하는 것이 필요하다.

본 연구에서는 계측장치에 발생하는 비용 없이 깊이별 토양온도 데이터를 습득하기 위한 방편으로 인공지능을 활용하여 시간에 따른 깊이별 토양온도를 예측하는 모델을 만들고 모델을 통해 도출된 예측값을 실측값과 비교하여 실제 농업에 적용이 가능한지 검토하였다.

RNN, LSTM, GRU 세 개의 머신러닝 모델과 지온에 영향력을 미칠 것이라 판단되었던 기온, 습도, 풍향, 풍속, 일사량 다섯 가지 기상인자를 활용하여 10cm, 20cm, 30cm 세 곳의 깊이에서의 시간 단위 지온 데이터를 정확하게 예측하는 최적의 매개변수와 모델 조합을 선정하였고, 시간에 따라 주기성을 갖는 월평균기온, 실시간 태양고도, 타지역 깊이별 토양온도 데이터 세 가지 주기성 인자를 추가하여 정확도 상승에 기여하는지 그 여부를 확인하였다. 그리고 가장 성능이 좋다고 평가된 두 개 조합으로부터 만들어진 예측값을 실측값과의 패턴분석과 잔차분석을 통해 유사성을 확인하여 실제 농업활동에 적용 가능성을 판단하였다. 평가는 Root Mean Square Error(RMSE), Nash - Sutcliffe Efficiency(NSE), Determination Coefficient( $R^2$ ) 세 가지 지표의 값을 통해 이루어졌다.

실험 결과 기온, 습도, 풍속, 일사량, 월평균기온 데이터와 LSTM 모델을 활용한 매개변수와 모델 조합이 가장 높은 성능을 보였고, 기온, 습도,

풍속, 일사량, 타지역의 깊이별 토양온도 데이터와 RNN 모델을 활용한 매개변수와 모델 조합이 다음으로 높은 성능을 보였으며, 실측값과의 잔차분석 결과 잔차의 정규성이 있으며 모든 깊이에서의 이상치율이 1% 내외로 나타나는 것을 확인하였다.

실험 과정을 통해 선별된 두 데이터세트, 모델 조합의 예측값 모두 평가 지표, 패턴분석, 잔차분석을 통해 실측값과 높은 일치성을 보인다고 판단되며, 실제 농업활동에 적용 가능하다고 평가하였다.

**주요어 : 토양온도, 밭, 인공지능, 시계열 예측**

**학번 : 2020-25115**

# 목 차

1. 서론 .....	1
1.1. 연구배경 .....	1
1.2. 문제점과 필요성 .....	4
1.3. 연구목적 .....	5
2. 문헌조사 .....	6
2.1. 용어와 관련 이론 .....	6
2.1.1 기상인자 매개변수 .....	6
2.1.1.1. 기온 .....	6
2.1.1.2. 습도 .....	6
2.1.1.3. 풍향 및 풍속 .....	7
2.1.1.4. 일사량 .....	8
2.1.2. 주기성을 나타내는 매개변수 .....	8
2.1.2.1. 전국 월평균기온 .....	9
2.1.2.2. 태양고도 .....	10
2.2. 지온의 특성 .....	11
2.3. 지온이 작물생육 미치는 영향 .....	16
2.4. 노지 스마트팜에서의 환경모니터링 연구 .....	18
2.5. 노지에서 깊이별 지온데이터 예측 연구 .....	21
2.6. 시계열데이터 예측을 위한 연구 .....	24
3. 자료 및 방법 .....	26
3.1. 공공기상자료 확보 .....	28

3.1.1. 공공기상대 데이터 .....	28
3.1.2. 학습용 데이터의 전처리 .....	29
3.2. 지온예측 평가용 지온데이터 .....	30
3.2.1. 외부기상대 및 토양센서 .....	30
3.3. 매개변수 선정 .....	33
3.4. 인공지능 모델 선정과 학습 .....	34
3.5. 깊이별 지온 예측 절차 및 평가 방법 .....	37
4. 결과 .....	38
4.1. 기상인자 데이터를 이용한 최적의 매개변수 선택	38
4.1.1. 소거를 이용한 기상인자 영향력 순서 평가 .....	38
4.1.2. 전진 선택법을 통한 기상인자 최적의 매개변수 조합 선택 .....	51
4.2. 주기성변수 .....	62
4.3. 데이터 비교 및 분석 .....	74
4.3.1. 추세 비교 .....	74
4.3.1.1. 10cm .....	75
4.3.1.2. 20cm .....	80
4.3.1.3. 30cm .....	85
4.3.2. 잔차 분석 .....	90
4.3.2.1. Q-Q Plot을 이용한 정규성 진단 .....	90
4.3.2.2. 표준화 잔차를 이용한 이상치 검출 .....	91
4.3.2.2.1. 예측값A .....	92
4.3.2.2.2. 예측값B .....	95
5. 결론 .....	97
6. 참고문헌 .....	100

## <List of Figures>

Figure 1 Changes in the ratio of the agricultural population and	1
Figure 2 Monthly average temperature data for the past 10 years.	9
Figure 3 Schematic diagram of the heat exchange mechanism (Sedighi et al., 2016)	12
Figure 4 The experimental methods of previous experiments and the experimental methods of this experiment	27
Figure 5 Data collected from public weather stations.	28
Figure 6 Satellite Image of Test Field	30
Figure 7 Picture of outside expectation (left) and soil sensor (right)	31
Figure 8 Location of external expectations and soil sensors(left)	32
Figure 9 RNN Model Structure(Colah's Blog, 2015)	34
Figure 10 LSTM Model Structure(Colah's Blog, 2015)	35
Figure 11 GRU Model Structure(Zhou, 2019)	36
Figure 12 Scatter plot graph of each set of RNN models at a	39
Figure 13 Scatter plot graph of each set of LSTM models at a	39
Figure 14 Scatter plot graph of each set of GRU models	40
Figure 15 Scatterplot graph of each set of RNN models	41
Figure 16 Scatterplot graph of each set of LSTM models	41
Figure 17 Scatterplot graph of each set of GRU models at	42
Figure 18 Scatterplot graph of each set of RNN models	43
Figure 19 Scatterplot graph of each set of LSTM models	43
Figure 20 Scatterplot graph of each set of GRU models	44
Figure 21 RMSE, NSE, and R2 values of RNN model predictions	45
Figure 22 RMSE, NSE, and R2 values of LSTM model predictions	45
Figure 23 RMSE, NSE, and R2 values of GRU model predictions	46



Figure 24 RMSE, NSE, and R2 values of RNN model predictions	····47
Figure 25 RMSE, NSE, and R2 values of LSTM model predictions	·47
Figure 26 RMSE, NSE, and R2 values of GRU model predictions	····48
Figure 27 RMSE, NSE, and R2 values of RNN model predictions	····49
Figure 28 RMSE, NSE, and R2 values of LSTM model predictions	·49
Figure 29 RMSE, NSE, and R2 values of GRU model predictions	····50
Figure 30 Scatterplot graph of each set of RNN, LSTM, GRU models	····52
Figure 31 Scatterplot graph of each set of RNN, LSTM, GRU models	····53
Figure 32 Scatterplot graph of each set of RNN, LSTM, GRU models	····54
Figure 33 RMSE, NSE, and R2 values of RNN model predictions	····55
Figure 34 RMSE, NSE, and R2 values of LSTM model predictions	·55
Figure 35 RMSE, NSE, and R2 values of LSTM model predictions	·56
Figure 36 RMSE, NSE, and R2 values of RNN model predictions	····57
Figure 37 RMSE, NSE, and R2 values of LSTM model predictions	·57
Figure 38 RMSE, NSE, and R2 values of GRU model predictions	····58
Figure 39 RMSE, NSE, and R2 values of RNN model predictions	····59
Figure 40 RMSE, NSE, and R2 values of LSTM model predictions	·59
Figure 41 RMSE, NSE, and R2 values of GRU model predictions	····60
Figure 42 Scatter plot graph of each set of RNN, LSTM, GRU models	····63
Figure 43 Scatter plot graph of each set of RNN, LSTM, GRU models	····64
Figure 44 Scatter plot graph of each set of RNN, LSTM, GRU models	····65
Figure 45 RMSE, NSE, and R2 values of RNN model predictions	····66
Figure 46 RMSE, NSE, and R2 values of LSTM model predictions	·66
Figure 47 RMSE, NSE, and R2 values of GRU model predictions	····67
Figure 48 RMSE, NSE, and R2 values of RNN model predictions	····68
Figure 49 RMSE, NSE, and R2 values of LSTM model predictions	··68
Figure 50 RMSE, NSE, and R2 values of GRU model predictions at a depth of 20 cm	····69
Figure 51 RMSE, NSE, and R2 values of RNN model predictions at a depth of 30 cm	····70

Figure 52 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 30 cm .....	70
Figure 53 RMSE, NSE, and R2 values of GRU model predictions at a depth of 30 cm .....	71
Figure 54 RMSE, NSE, and R2 values for each data set and model combination at a depth of 10 cm. ....	72
Figure 55 RMSE, NSE, and R2 values for each data set and model combination at a depth of 20 cm. ....	73
Figure 56 RMSE, NSE, and R2 values for each data set and model combination at a depth of 30 cm. ....	73
Figure 57 Graph of annual changes in measured values .....	75
Figure 58 Graph of annual change of predicted value A .....	76
Figure 59 Graph of annual change of predicted value B .....	77
Figure 60 Daily variation graph of the measured value, predicted value A, and predicted value B at August 15 at a depth of 10 cm. ....	79
Figure 61 Graph of annual changes in measured values .....	80
Figure 62 Graph of annual change of predicted value A .....	81
Figure 63 Graph of annual change of predicted value B .....	82
Figure 64 Daily variation graph of the measured value, predicted value A, and predicted value B at August 15 at a depth of 20 cm. ....	84
Figure 65 Graph of annual changes in measured values .....	85
Figure 66 Graph of annual change of predicted value A .....	86
Figure 67 Graph of annual change of predicted value B .....	87
Figure 68 Daily variation graph of the measured value, predicted value A, and predicted value B at August 15 at a depth of 30 cm. ....	89
Figure 69 Q-Q plot for each depth of predicted values A and B .....	91
Figure 70 Time zone and standardized residuals value of outliers .....	92
Figure 71 Time zone and standardized residuals value of outliers by depth of predicted value A. ....	93
Figure 72 Time zone and standardized residuals value of outliers by	

depth of predicted value B. .... 95

## <List of Tables>

Table 1 Purpose, Network, measurement depth indicated in the reference .....	20
Table 2 latitude, longitude, altitude above sea level, and climatic characteristics of each public weather station. ....	29
Table 3 Specifications of each weather and soil sensor. ....	31
Table 4 The parameters used in the experiment .....	33
Table 5 Combination of parameters of each dataset .....	38
Table 6 Parameters of each data set .....	51
Table 7 Parameters of each data set .....	62
Table 8 Comparison table of annual best records, lowest records, and average values of measured values, predicted values A, and predicted values B .....	78
Table 9 Comparison table of the daily highest, lowest, and average values of August 15 of the measured value, predicted value A, and predicted value B .....	79
Table 10 Comparison table of annual best records, lowest records, and average values of measured values, predicted values A, and predicted values B .....	83
Table 11 Comparison table of the daily highest, lowest, and average values of August 15 of the measured value, predicted value A, and predicted value B .....	84
Table 12 Comparison table of annual best records, lowest records, and average values of measured values, predicted values A, and predicted values B .....	88
Table 13 Comparison table of the daily highest, lowest, and average values of August 15 of the measured value, predicted value A, and	

predicted value B ..... 89

# 1. 서론

## 1.1. 연구배경

현재 농촌은 인구감소 및 고령화로 많은 어려움을 겪고 있다. 2021년도에 갱신된 통계청의 ‘농가 및 농가 인구’ 자료에 따르면 최근 10년간 농업 인구는 감소하는 추세를 보이고 있는 반면, 65세 이상 고령인구의 비율은 증가하는 추세를 보이고 있다. 그 결과 노동 부족 문제로 인해 지역 소멸과 식량안보를 위협받고 있다.

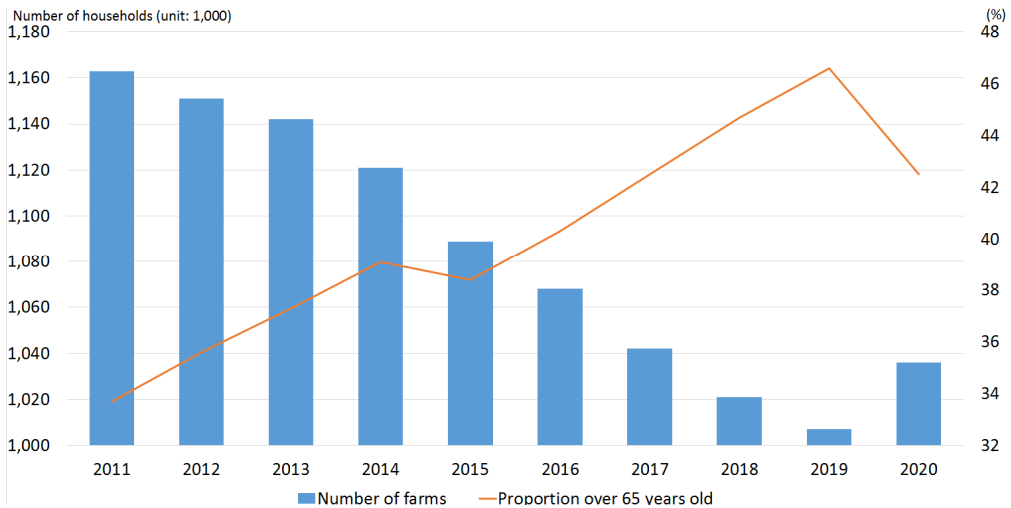


Figure 1 Changes in the ratio of the agricultural population and the elderly population over the past 10 years(Statistics Korea, 2021)

이러한 위기를 극복하기 위해 정부 해결책으로 농업에 ICT(Information and Communication Technology) 기술을 적용한 스마트 농업에 관한 기술을 개발하고 적용 및 보급하는 중이다(Youm et al., 2018). 스마트 농업정책에 의하여 각종 스마트한 시설과 장비에 데이터와 센서가 결합되어 농업의 첨단 산업화가 이어지고 있지만, 우리나라의 경우 통계청의 ‘2020년 경

지면적 조사 결과' 자료에 따르면 현재 노지 면적이 1,565,000ha로 농경지 전체 면적의 95% 이상을 차지하고 있음에도 불구하고 데이터를 기반으로 한 디지털 농업의 기술은 시설원예와 시설축산을 중심으로 개발되어왔다. 정부가 지원하는 노지에서의 스마트 농업은 시범사업은 주로 농업용 작업기의 자동화, 무인화에 집중되어 왔기 때문이다(Jo et al., 2021).

생산 비용을 줄이면서 더 많은 작물을 생산하기 위해 각종 센서, GPS(Global Positioning System), IOT(Internet of Things) 등을 활용하여 기상, 토양, 작물 정보를 분석하고 자원의 사용을 최소화하며 생산성을 극대화하는 정밀농업 기술이 노지에서도 적용되어야 할 필요가 있다. 이중 토양은 농작물 생장에 직접적으로 관여하므로 작물의 정상적 성장을 위해서는 필수적이기 때문에 깊이에 따른 토양 환경을 계측하고 관리하여 알맞은 농작 활동을 하는 것이 중요하다. 토양온도의 경우 작물은 종자의 발아에서부터 영양생장, 저장 기관의 발달, 개화, 성숙, 노화에 이르기까지 온도에 따라 다양한 생육 반응을 보이며, 식물의 뿌리로 흡수되는 비료는 지온의 영향을 받기 때문에 시간에 따른 변화량을 확인하는 것이 중요하다. 현재 농업현장에서 비료 시비 결정과 관개 제어 시스템 개발을 위해 토양온도 데이터를 활용 중이다. 비료 시비의 경우 근권부의 토양온도가 30℃ 이상 되면 인산칼륨의 흡수 상태가 점점 나빠지게 되며 반대로 낮아질 경우에는 질소의 흡수가 감소하게 되기 때문에 적정 수준의 온도 관리가 필요하다. 토양온도는 비종의 선택에도 영향을 끼치는데 인산의 경우 근권부의 온도가 높을 때에는 흡수가 잘되나 낮은 경우에는 흡수가 잘 안되는 경향으로 지온이 낮은 시기에는 과린산 석회의 시용이 바람직하다.

그러나 토양 환경을 실시간을 분석함에 있어 외부기상대와 더불어 토양 계측 센서와 데이터를 실시간으로 확인할 수 있는 모니터링 시스템은 농작 활동을 하는 농업인에게 경제적으로 부담이 되는 가격으로 시중에 제공되고 있다. 시간에 따른 깊이별 지온의 변화량은 기상, 지형, 토양성분 등 주변 환경 인자의 영향을 받는 것으로 알려져 있으며, 그중 기상환경은 토양에 직접적으로 영향을 미치는 인자이기 때문에 기상인자와 인공지능 모델을 통해 깊이별 지온을 예측하는 몇몇 연구들이 진행되었다.

인공지능 머신러닝 모델 중 하나인 RNN(Recurrent neural network) 모

델은 유닛 간의 연결이 순환적 구조를 갖는 특징을 갖고 있는 모델이다. 몇 단계 전으로 거슬러 올라가 데이터를 반영하는 신경망 알고리즘이기 때문에 1회의 데이터가 아닌 과거의 데이터도 학습에 활용이 가능하며, 이러한 특징 때문에 시간의 흐름에 따라 변화하는 시계열 데이터를 처리하는데 좋은 성과를 보인다(Lee et al., 2021). 그러나 RNN 모델은 입력 데이터가 길어질수록 거리가 떨어진 데이터의 정보를 삭제하는 장기 의존성(Long-Term Dependency)문제를 가지고 있다. 이러한 장기 의존성 문제를 해결하기 위해 LSTM(Long Short-Term Memory) 모델이 개발되었고, 처리 시간이 길다는 LSTM 모델의 단점을 보완하기 위해 학습 속도가 빠르고 LSTM 모델과 비슷한 성능을 보이는 GRU(Gated recurrent unit) 모델이 개발되었다. 현재 많은 연구에서 시계열 데이터를 분석하고 예측 위해 RNN 모델과 RNN 모델을 기반으로 개발된 LSTM 모델, GRU 모델이 사용 중에 있는 반면, 시간에 따라 변화량을 갖는 시계열 데이터인 깊이별 지온 데이터를 예측하는 연구에서는 사용이 보고된 바 없다.

따라서 본 연구에서는 RNN, LSTM, GRU 세 개의 모델과 지온에 영향력을 미칠 것이라 판단되었던 기온, 습도, 풍향, 풍속, 일사량 다섯 가지 기상인자 데이터를 통해 10cm, 20cm, 30cm 세 곳의 깊이에서의 시간 단위 지온 데이터를 예측하고 정확성을 평가하여 농업에 적용 가능성을 판단하였다.



## 1.2. 문제점과 필요성

작물 재배에 있어서 토양은 절대적으로 중요한 요소이며, 토양의 온도가 작물의 발아는 물론 생육과 수량에도 영향을 미친다는 것은 잘 알려져 있다. 관행 농업에서는 절기에 따른 파종과 파종 이후 생육 단계를 확인하면서 시비하고, 적정 토양온도를 높이는 방법도 멀칭이나 두둑을 이용하는 방법에 불과하였다. 토양온도가 수량과 생육에 미치는 영향에 대한 연구가 많지만 토양온도를 어떻게 조절할 수 있는지 공학적인 해법은 부족한 상황이다.

한편 스마트 농업이 발달함에 따라 각종 센서에서 얻은 정보를 바탕으로 농작업의 시기와 방법, 정도를 조정하는 정밀농업의 개념이 등장하였으나 작물의 생장에 중요한 토양온도에 대한 연구는 매우 부족하다. 노지 농업에서 이용되는 환경정보는 과수원을 중심으로 고부가 가치 작물 재배에 기상타워를 이용한 기온과 습도, 일조량, 강우량이 있으며 일부 연구자들은 토양의 수분함량과 pH를 활용하고 있으며 깊이별 토양온도를 이용하는 연구는 매우 제한적이다. 토양온도에 대한 정보는 기상청이 운영하는 지역 기상대에서 제공되고 있으나 이 정보를 농업에 활용하는 방안에 대한 연구는 찾기 어렵다.

스마트 노지 농업에 적용되는 센서가 제한적인 이유는 토양의 온도나 습도가 국부적인 토양의 조성에 크게 영향을 받아 같은 농경지 안에서도 변이가 크며 센서들의 가격이 높고, 농경지 내부에 설치하므로 농작업에 불편을 주는 등 여러 가지 원인을 생각할 수 있으나 데이터에 입각한 작물 재배를 지향하는 digital agriculture 기술에서 보면 발아와 생육, 수량에 밀접한 토양의 깊이별 온도 변화는 중요한 연구 대상이다. 토양온도에 영향을 미치는 요인은 매우 다양하며 이론적으로 예측하는 것이 어려운데

따라서 농업인들이 농경지 내에 설치하기를 부담스러워하는 토양온도센서 대신에 전국 기상대에서 제공하는 지중온도를 이용하여 임의 지역 농경지의 토양온도를 예측하는 기술을 개발하고, 이 정보에 근거하여 노지 농업을 스마트하게 수행할 수 있는지를 판단할 필요가 있다.

### 1.3. 연구목적

본 연구의 목적은 전국의 공공기상대에서 제공하는 기상인자와 깊이별 토양온도 데이터를 이용하여 시간에 따른 토양온도를 예측하는 인공지능 모델을 개발하고 농업에 적용할 수 있는지 평가하기 위한 것이다.

본 연구에서 개발된 인공지능 모델을 평가하기 위하여 농업활동이 이루어지고 있는 특정 밭 지점에서 11개월간 수집한 기상인자와 깊이별 토양온도 데이터를 확보하여 이를 활용하여 인공지능 모델의 예측 성능을 검토하고자 하였으며 구체적인 실험과정은 다음과 같다.

- 1) 토양의 깊이별 온도를 예측하는데 유효한 영향을 미치는 기상인자를 매개변수로 선정하고 각 변수가 지온 예측에 정확도를 높이는 데 미치는 영향력의 여부와 그 크기를 평가하여 최적의 기상인자를 검출한다.
- 2) 토양의 깊이별 온도의 연간, 일간 패턴 변화를 읽고 적용하여 변화량을 예측할 수 있도록 주기성 인자를 포함시켜 정확도의 상승 여부를 확인한다. 주기성 인자는 주기성을 잘 설명하는 인자를 선정한다.
- 3) 최적의 예측 성능을 보이는 것으로 사료되었던 데이터 세트, 러닝 머신의 조합을 평가지표를 통해 두 개를 선정한다. 그리고 두 개의 조합을 취득한 예측값과 실측값의 유사성을 확인하여 실제 농업활동에 적용이 가능한지 평가한다. 유사성은 다음을 통해 확인한다.
  - 연간 추세와 일별 패턴의 유사성을 확인한다.
  - 그리고 잔차분석을 통해 정규성 여부를 확인하고, 이상치를 검출하여 회귀 모델의 적합성을 확인한다.

## 2. 문헌조사

### 2.1. 용어와 관련 이론

#### 2.1.1. 기상인자 매개변수

##### 2.1.1.1. 기온

기온은 토양 온도의 예측에 있어 예측값에 영향을 미치는 중요한 기상인자 중 하나이다. 지표면은 대기와 현열 이동을 통해 높은 곳에서 낮은 곳으로 열이 전도되는 열 교환을 하며, 지면에 닿아 가열된 태양 복사열은 전도를 통해지면 위에 위치한 공기층을 가열하여 기온을 상승시킨다. 지구 온난화는 대기 온도뿐 아니라 토양온도와 강수 패턴에도 영향을 미칠 수 있다는 것은 기온과 지온은 상관관계가 크다는 것을 입증한다 (Jungqvist et al., 2014).

기상인자를 활용하여 깊이별 지온을 예측한 모든 연구에서 기온은 지온에 가장 영향력을 미치는 기상인자로 확인되었으며, 기상인자와 인공지능을 활용해 깊이별 지온을 예측했던 Alizamir et al.(2019)의 연구에서는 50cm 깊이까지는 기온을 단일 매개변수로 활용한 예측의 정확도가 가장 높다는 결과가 도출되었다. 따라서 본 연구에서도 깊이별 지온을 예측하는데 있어 시간에 따른 지온 값을 매개변수로 활용하였다.

##### 2.1.1.2. 습도

토양 내의 수분함량은 지온에 영향을 미친다. 토양은 토양입자와 유기물로 구성된 고상, 토양 내 수분을 의미하는 액상, 토양 내 공기를 의미하는 기상으로 이루어져 있는데, 물의 열전도율은 공기의 30배에 달하기 때문에 토양 내 액상의 비율이 높을수록 열전도가 빠르다. 또한 수분이 지하 내에서 이동할 때 열을 운반하기 때문에 토양 내 열확산에 영향을 미친다.

습도의 경우 토양 내의 수분함량과 밀접한 관련이 있다. 토양 내의 수분함량이 높을수록 증발량이 증가해 습도가 증가하며, 대기 중의 수증기는 액화되어 지면으로 침투한다. 뿐만 아니라 습도는 evaporation flux에 영향을 미치는데, evaporation flux  $E$  는 Eq. (1)을 통해 구할 수 있다 (Sedighi et al., 2016).

$$E = \rho_a r_a (q - q_{air}) \quad (1)$$

$\rho_a$  : air density

$r_a$  : aerodynamic resistance

$q$  : specific humidity of the soil at the ground surface

$q_{air}$  : humidity of air

기상인자를 활용하여 깊이별 지온을 예측했던 대부분의 기존 연구들에서 습도는 예측 정확도를 높이는 데 있어 중요한 매개변수로 활용되었다. 따라서 본 연구에서는 보다 높은 정확도로 지온을 예측하기 위해 시간에 따른 습도 데이터를 매개변수로 활용하였다. 습도는 상대습도(%)를 사용하였다.

### 2.1.1.3. 풍향 및 풍속

바람을 통한 증발은 토양의 온도와 습도에 영향을 미친다. 지표면에 바람이 지나갈 때 지표면에서 마찰항력과 난류를 생성시키며 현열과 잠열이 상승될 수 있다(Dingman, 1994). Energy balance equation에 따르면 현열 복사의 크기는 공기역학저항(aerodynamic resistance)과 음의 상관관계를 갖는데, 공기역학저항  $r_a$ 은 Eq. (2)를 통해 구할 수 있다(Sedighi et al., 2016).

$$r_a = \frac{k^2 u_z}{\left[\ln\left(\frac{z}{z_0}\right)\right]^2} \quad (2)$$

$k$  : von Karman constant

$u_z$  : wind speed

$z_0$  : surface roughness,

Eq. (2)를 통해 풍속은 현열 복사의 크기를 결정하는 데 영향을 미치는 기상 인자라는 것을 유추할 수 있기 때문에 본 연구에서는 기상인자 매개변수로써 풍속과 풍속이 가해지는 방향을 나타내는 풍향을 활용하였다.

#### 2.1.1.4. 일사량

태양 에너지는 행성 주위에 분산되어 있다. 지구로 유입되는 태양에너지의 약 54%는 우주로 반사되거나 대기에 흡수되며, 나머지 46%는 지표면에 흡수된다(Bilgili, 2010). 이렇게 흡수된 에너지는 지표면에서 흡수와 방출, 대기와의 사이에 잠열과 현열 수수가 이루어진다. 태양복사 에너지는 토양 표면과 내부에 있는 열에너지의 원천이 되는 에너지이다. 따라서 본 연구에서는 일사량 데이터를 매개변수로써 활용하였다.

#### 2.1.2. 주기성을 나타내는 매개변수

토양온도는 기후변화뿐 아니라 연간 주기와 주행성에도 영향을 받는다. (Kang et al., 2000). 다양한 깊이에서 토양온도의 연간 변화는 사인과 함수를 사용하여 추정할 수 있으며, 토양 온도의 연간 주기는 8월에 최고치를 기록하고, 12월과 1월 사이에 최저치를 가진다(Bilgili, 2010).

따라서 본 연구에서는 기상변수 이외에 계절성 및 시간의 변화량을 나타낼 수 있는 주기성 변수가 모델의 예측 정확도를 높이는데 기여할 수 있

는지 확인하였다.

전국 월평균기온, 해당 지역의 태양고도, 특정 지역의 깊이별 지온 데이터 세 가지 변수를 각각 기상변수 외의 변수로 활용하여 정확도의 상승 여부를 확인하였다.

#### 2.1.2.1. 전국 월평균기온

Milinkovitch cycle에 따르면 지구의 자전축은 약 26,000년마다 한바퀴씩 세차운동을 하며 이러한 과정을 통해 지구 자전축의 경사는 41,000년을 주기로 21.5°에서 24.5° 사이를 오르내린다.

한반도의 경우 기온의 변동은 1년 주기로 월별 또는 계절별로 특성을 갖고 있다. 겨울에 변동이 가장 심하고 여름은 변동이 가장 적다. 봄과 가을은 그 중간정도 된다.

Shin et al.(2009)은 CSEOF 분석기법을 통해 한반도 기온은 4~5년 주기로 변동하며, 1987~1998년 사이 강도가 약 0.5℃ 약해지고 1999년 이후 다시 강해지는 추세를 보이는 것을 확인하였다.

본 연구에서는 연별, 일별 주기성을 갖는 기온을 주기성 지표로 활용하기 위해 ‘전국 월평균기온’을 매개변수로 활용하였다.

‘기상청’의 자료에 따르면 해당 매개변수의 경우 최근 10년간 연간 주기성을 확인할 수 있다.

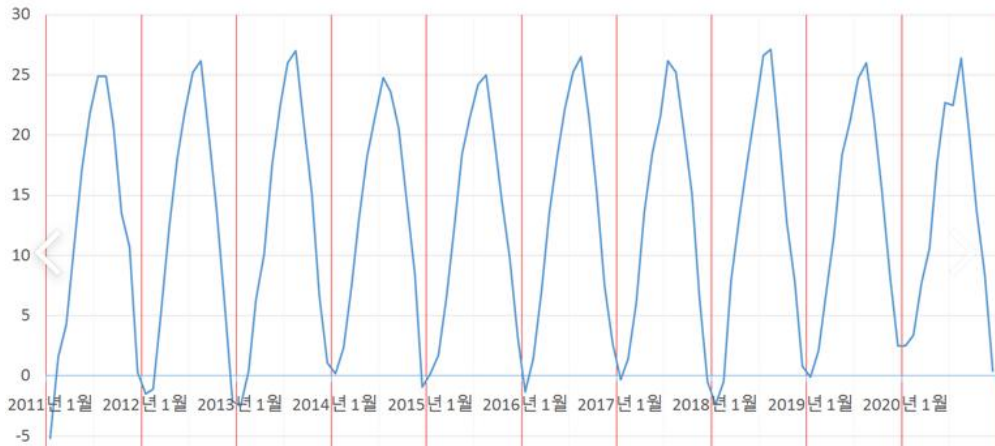


Figure 2 Monthly average temperature data for the past 10 years.

### 2.1.2.2. 태양고도

태양의 고도는 일출 후 상승하여 낮 12에 가장 높고, 그 이후 다시 낮아진다. 같은 시각 태양의 고도는 계절에 따라 달라지는데 지구가 태양 주변을 공전할 때 자전축이 기울어진 상태로 공전을 하기 때문이다. 따라서 본 연구에서는 연주기를 가지고 계절과 시간에 따른 변화량을 가지고 있는 태양고도를 주기성 변수로 활용하여 모델을 만들고 정확성을 평가하였다.

태양 고도의 경우 해당 지역의 위도, 경도 그리고 시간을 통해서 구할 수 있다. 먼저 균시차(Equation of Time)를 알아야 되는데 균시차란 태양의 위치로 확인된 진태양시와 일상생활에서 사용하는 평균태양시의 차이를 나타내며 균시차  $EOT$ 는 Eq. (3)을 통해 구할 수 있다.

$$EOT = 229.2 \times (0.000075 + 0.001868 \times \cos(B) - 0.032077 \times \sin(B) - 0.014615 \times \cos(2B) - 0.04089 \times \sin(2B)) \quad (3)$$

$$B = (n - 1) \frac{360}{365}$$

$n$  : day of year

그 다음 시간각(Hour Angle)을 구해야 한다. 시간각이란 자오면 (meridian plane)과 지축과 태양의 위치를 연결하는 면이 이루는 각도를 의미하며, 시간각  $\omega$  는 Eq. (4)를 통해 구할 수 있다.

$$\omega = \frac{LT \times 60 + 4 \times (LL - LST) + EOT}{60} \times 15 - 180 \quad (4)$$

$LT$  : local standard time

$LL$  : local longitude

$LOT$  : meridian longitude

그 후 태양적위(Solar Declination)을 구해야 되는데 태양적위란 태양의 방향과 북극의 방향이 이루는 각의 여각이다. 태양적위  $\delta$  는 Eq. (5)를 통해 구할 수 있다.

$$\delta = 23.45 \times \sin\left[\frac{360}{365} \times (284 + n)\right] \quad (5)$$

이렇게 태양적위의 값까지 계산하면 태양 고도를 구할 수 있다. 태양고도는 Eq. (6)을 통해 구할 수 있다.

$$\sin(\beta) = \cos(L)\cos(\delta)\cos(\omega) + \sin(L)\sin(\delta) \quad (6)$$



## 2.2. 지온의 특성

기상학에서 지표온도는 지면온도와 지중온도로 구별되며 지면온도는 맨 땅 또는 짧은 잔디 밑 지면의 온도를 말하며, 지중 온도는 통상적으로 0.05, 0.1, 0.2, 0.3, 0.5, 1, 1.5, 3, 5 m 깊이의 토양 온도를 의미한다. 지표 온도는 지구온난화로 인하여 연평균 값이 최근에 지속적으로 증가하고 있다. 한편 농업생명과학에서 지표는 토양(soil)로 정의되며 토양은 식물을 지지하고 수분과 영양분을 공급하며 식물의 지하부인 뿌리의 환경을 제공하며 뿌리가 자라는 지중 환경을 근권부(root zone)라고 한다.

식물은 근권부에서 발아하고, 근권부에서 영양을 공급받으며, 근권부의 수분을 흡수하며, 이러한 생명현상은 식물의 효소와 촉매의 영향을 받으므로 근권부 토양온도는 화학적, 생리학적 과정의 속도에 영향을 미치는 주요 요인이다(Sauer et al., 2005).

근권부 토양 온도는 접하고 있는 대기권과 열평형을 이루는 과정에서 결정되는 물리적 성질로서 국부적인 열 수지(heat balance)로 설명될 수 있다. 토양에 제공되는 원천 에너지는 태양의 복사에너지이며 Sedighi et al.(2016)는 지표면의 복사를 단파 열 복사, 장파 열 복사, 현열 복사, 잠열 복사 네 가지 주요 메커니즘으로 Figure 4와 같이 설명하면서 열수지 방정식을 Eq. (7)과 같이 제안하였다.

$$H = H_{SW}^{Absorbed} - (H_{LW}^{Net} + H_{SN} + H_{\leq}) \quad (7)$$

$H$  : total radiation heat flux absorbed or emitted at the soil surface

$H_{SW}^{Absorbed}$  : adsorbed short wave radiation flux

$H_{SW}^{Net}$  : net long-wave radiation flux

$H_{SN}$  : sensible heat flux

$H_{\leq}$  : latent heat flux

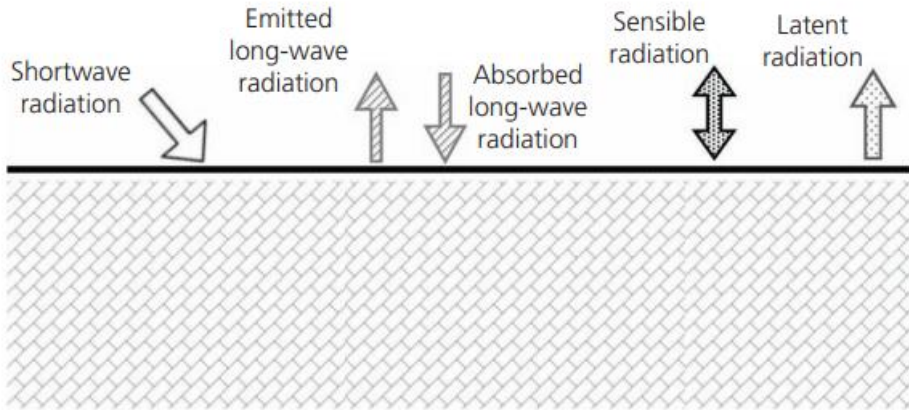


Figure 3 Schematic diagram of the heat exchange mechanism(Sedighi et al., 2016)

단파 복사(short wave radiation)는 지구 표면에 복사되는 태양에너지의 조합이다. 지구 표면에 도달하는 총 단파복사는 분석하고자 하는 지점의 위도와 경도, 해당 지점의 평균적인 구름의 양, 분석하는 시기, 지표면의 피복 조건과 반사율 등에 따라 달라지며 연간 일정한 주기로 변화하는데, Deardorff et al.(1978)는 그중 표면에 흡수되는 shortwave radiation flux의 비율을 Eq. (8)과 같이 제안하였다.

$$H_{SW}^{Absorbed} = \varepsilon_{SW} H_{SW} \quad (8)$$

$H_{SW}^{Absorbed}$  : heat flux associated with the absorbed shortwave radiation

$\varepsilon_{SW}$  : shortwave reflection factor associated with the ground surface type

장파 복사(long-wave radiation)의 경우 지표에서 방출되는 장파 복사와 대기에서 흡수되는 장파 복사의 조합으로 설명할 수 있는데 장파복사  $H_{SW}^{Net}$  는 Eq. (9)와 같다.

$$H_{LW}^{Net} = H_{SW}^{Absorbed} - H_{LW}^{Emitted} \quad (9)$$

$H_{SW}^{Absorbed}$  : radiation flux absorbed at the ground surface

$H_{LW}^{Emitted}$  : radiation flux emitted from the ground surface

물체에 의해 방출되는 에너지는 절대 온도의 네 번째 힘에 정비례한다는 Stefan - Boltzmann's law을 적용하면 Eq. (10)의 식을 얻을 수 있다 (Lewis et al., 2004).

$$H_{LW}^{Emitted} = \varepsilon_{LW}\varphi T^4 \quad (10)$$

한편 깊이별 토양 온도는 지표면에서 열전도에 의해 토양 내부로 전달되어 변화되기 때문에 열용량이 큰 토양의 특성 때문에 지표면의 온도 변화는 내부로 전달될수록 감소하고, 시간이 걸린다. 이러한 이유로 지온은 깊이에 따라 변화 폭이 달라지고 기온과 증감의 시간차가 존재한다(Bilgili, 2010). 그뿐만 아니라 지표면의 알베도, 수분함량 및 질감, 지형 변수, 잎의 면적과 같은 기상학적 이외의 환경요인도 영향을 미치는데(Bilgili, 2010), 겨울철에는 눈 덮힘 현상으로 기후와 토양이 분리되었을 때 눈 덮힘(snowpack)의 시기와 지속시간에 따라 지온에 미치는 기후의 영향력은 감소하게 된다.

현열과 잠열은 전도(conduction)과 대류(convection)에 의한 열이며 수분밸런스(moisture balance)와 함께 고려된다. 현열은 Eq. (11), 잠열은 Eq. (12)으로 표시할 수 있는데, 잠열을 결정하는 수분증발량은 강우량, 지표유실량, 지표침투량, 토양의 수분함량과 수분 평형을 이루는 상황에서 예측되는 물리량이다.

$$H_{SN} = \frac{\rho_a C_p}{r_a} (T - T_{air}) \quad (11)$$

$\rho_a$  : air density

$C_p$  : ground specific heat capacity

$r_a$  : aerodynamic resistance

잠열 복사는 열역학적 시스템에서 에너지의 흡수 또는 방출로 정의된다. 지표면에서의 수분증발과 관련이 많다. Sedighi et al.(2016)은 잠열복사  $H_{\leq}$  를 Eq. (12) 와 같이 제안 하였다.

$$H_{\leq} = L \times E \quad (12)$$

$L$  : latent heat of vaporisation

$E$  : evaporation flux

토양의 내부 영역을 통과하는 시간 단위의 열에너지 양은 토양의 열 유속으로 정의할 수 있다. 토양 내 열 흐름은 Fourier's Law가 적용되는 고체의 열류와 유사하다고 간주될 수 있는데,  $Z$  깊이에서 토양의 온도가  $T$  일 때 Eq. (13) 과 같이 나타낼 수 있다.

$$G = -\lambda \left( \frac{dT}{dZ} \right) = -\lambda \frac{T_{Z_1} - T_{Z_2}}{Z_1 - Z_2} \quad (13)$$

$G$  : heat flux density

$\lambda$  : thermal conductivity of the soil

$\frac{\sigma T}{\sigma Z}$  : vertical temperature gradient

Enrique et al. (1999)는 실험을 통해 시간  $t$ 가 경과되었을 때 깊이  $z$ 에서의 토양온도를 추정하는 식을 Eq. (14) 와 같이 제안 하였다.

$$T(z,t) = T_a + A_0 \exp\left(-\frac{z}{d}\right) \sin\left[\frac{2\pi(t-t_0)}{365} - \frac{z}{d} - \frac{\pi}{2}\right] \quad (14)$$

$T_a$  : constant groundtemperature

$A_0$  : annual amplitude of the surface soiltemperature

$t_0$  : lag time from arbitrary start date to the occurrence of the

minimum soil temperature in a year  
 $d$  : damping depth

토양의 열전도율은 고상분율, 수분함량, 부피밀도의 구성에 따라 달라진다(Sauer et al., 2005). 따라서 지중 온도는 국부적인 토양의 조건(고상분율, 부피밀도, 수분함량, 등)과 위치 조건(위도와 경도 및 고도, 평균적인 운량, 경사도, 등), 분석 시기(월과 일), 기상조건 (기온, 상대습도, 강우량, 일조량) 등 많은 환경인자의 영향을 받음을 알 수 있다.

### 2.3. 지온이 작물생육 미치는 영향

모든 식물은 발아, 영양생장, 개화, 생식생장의 단계를 거치며 이 과정에서 적절한 온도와 적절한 수분, 적당한 산소가 필수적이다. 토양은 식물이 발아되는 환경이며, 식물이 줄기와 잎, 뿌리로 분화하여 성장하는 과정에서 수분과 양분을 공급하는 기반이 된다.

종자의 발아 조건은 수분, 온도, 산소 외에도 빛이 필요하며 토양 환경이 이 모든 조건을 결정한다. 발아는 생리적 현상으로 수분을 흡수한 종자에서 저장물질(貯藏物質)의 전화(轉化)·전류(轉流)·호흡작용, 그 밖의 생리적 기능을 일으키며 배(胚)가 종피를 뚫고 나와 줄기와 뿌리를 형성한다. 이러한 생리적 현상은 효소와 촉매에 의해 진행되므로 적절한 온도가 제공되지 않으면 발아가 실패하거나 지연된다.

작물은 품종마다 최대 성장을 위한 최적의 온도 범위를 가지고 있다(Tenge et al., 1998). 밭에서 자라는 작물들의 경우 고구마는 밭에 묘를 심은 후 뿌리를 내리는데 15℃ 이상의 온도가 필요하며, 최적의 지온은 30℃이다. 묘를 심은 후 지온이 15℃ 이상 유지되지 않으면 고구마 수가 감소하기 때문에 적절한 지온유지가 필요하며(RDA, 2021), 양과는 파종 후 발아하는데 걸리는 시간이나 발아율이 온도에 따라 달라진다(RDA, 2021). 옥수수는 토양온도에 따라 발아 소요시간이 달라지는데 13℃에서는 18~25일, 15~18℃에서는 8~10일, 21℃ 이상 되면 4~6 일이 소요된다(RDA, 2021).

발아 이후에도 식물의 뿌리성장은 토양의 온도에 영향을 받는다(Kaspa, 1992). 토양 온도는 분해와 무기화작용 같은 생물학적 과정을 통해 뿌리 표면의 영양소의 가용성을 변화시킨다(Bassirirad, 2000). 토양 온도의 증가는 뿌리 세포의 대사 활성의 증가와 측면 뿌리의 발달로 인해 뿌리 성장을 개선한다(Repo et al., 2004). 낮은 토양 온도는 조직 영양소의 농도를 감소시켜 뿌리 성장을 감소시킨다(Lahti et al., 2002).

Clarke et al.(2015)는 봄철 토양 온도, 탄수화물 비축량 동원, 영양소 섭취 및 질소 분할 사이의 연관성을 Shiraz 포도넝쿨에서 조사하여 이러한

무생물 인자가 어떻게 뿌리와 캐노피 발달을 변화시킬 수 있는지 확인하는 연구를 진행하였고 토양온도는 뿌리성장 역학을 변화시킬 수 있었고 작물의 지하와 지상간의 탄소 및 양분분배에 영향을 미친다는 것을 확인하였다.

Yun et al.(2012)는 가을철 배추재배에 있어서 비닐멀칭의 색상이 토양 온도 변화와 수량에 미치는 영향을 연구한 결과 색상별로 지온에 미치는 영향이 다르며 흑색비닐을 사용한 경우에 시비질소의 이용율이 가장 높았고 수량도 가장 많았다고 보고하였다.

이와 같이 토양의 온도는 발아율, 발아기간, 발아세력과 아주 밀접한 관계를 갖고 있을 뿐만 아니라 발아 이후의 작물생육은 물론 수량에도 직접 영향을 미치고 있음을 알 수 있다.

## 2.4. 노지 스마트팜에서의 환경모니터링 연구

노지는 시설원예와 달리 환경을 제어하는 데 한계가 분명하기 때문에 외부환경인자를 이용한 데이터 농업의 기술발전이 제한적이며, 특히 2010년대 이후에 밭 농업기계화에 주력하다가 2010년대 후반에 이르러 노지에서의 스마트 팜과 자율주행에 대한 연구가 시작된 우리나라의 연구 상황 때문에 제대로 이루어지지 않았다.

해외의 연구사례를 보면 Haggard et al.(2010)는 Transylvanian Plain 지역에 있는 토양의 온도에 영향을 미치는 인자를 찾아 토양온도상(soil temperature regimes)을 확인하고 이것을 바탕으로 지역 작물의 종자 발아를 위한 최적의 파종 일을 권고하기 위해 토양온도 모니터링 시스템을 개발하였다. 토양온도는 실험에 사용된 작물인 옥수수의 생장온도 일수(growing degree days)를 계산하는 데 사용되었으며, HOBO 사의 S-TMB-M002 온도센서를 이용하여 측정되었다.

Chen et al.(2014)은 다양한 깊이에서의 토양 환경을 파악하여 생장기 작물 뿌리 주변의 수분 수요를 파악하기 위해 CC2530 모듈과 WSN(Wireless Sensor Network)를 이용하여 다층 토양 온도와 습도에 대한 모니터링 시스템을 설계하였다. 개발된 시스템은 모니터링 노드, 게이트웨이 노드 및 명령을 통해 제어하는 컴퓨터 시스템 플랫폼 세 구성으로 구성되었다. 실시간으로 토양 온도와 습도 데이터를 파악하기 위해 10 cm, 20 cm, 30 cm, 40 cm 토양층 4개의 토양온도 및 수분함량의 변화량을 모니터링하였다.

Goap et al.(2018)는 물 소비량을 줄일 수 있는 효율적인 지능형 관개 시스템을 개발하는 연구를 하였다. 연구에서 실시간 모니터링 된 토양 온도는 기온, 자외선 복사, 습도 데이터와 SVR(Support Vector Regression) 모델과 토양 내 수분의 증발량을 추정하여 토양 내 수분함량을 예측하는 데 사용되었으며, VH-400 센서를 통해 계측되었다. 통신은 Zigbee를 활용하였다.

국내의 연구사례를 살펴보면 Lee(2011)는 USN(Ubiquitous Sensor Network)



기술을 이용한 농작물 모니터링 시스템의 해외적용 가능성을 확인하기 위해 미국 Napa지역에서 와인용 포도품종의 생장관리 시스템을 개발하는 연구를 하였다. 개발된 시스템은 기상 및 토양 데이터를 실시간으로 모니터링 한다. 이 중 토양온도는 토양의 습도와 함께 작물의 증산율을 계산하는데 사용되며, 증산율을 통해 관수량을 조절한다.

Youm et al.(2018)은 노지에서 농작물 생산성 향상과 여러 가지 토양오염 방지를 위한 방안으로 농작물 재배의 자동화, 무인화 및 지능화 실현을 위해 Lora 통신을 이용하여 넓은 노지에서 재배, 생산량, 판매실적을 관리하는 시스템을 개발하였다. IOT(Internet of Things)를 이용한 노지작물 재배 시스템에서의 환경 변수는 대기의 온도와 습도 외에, 토양의 온도, 함수율 전기전도도, 산성도였으며 이를 이용하여 시비량과 관수량을 제어하였다. 산성화된 토양에 밭작물 재배에 적당한 산도 6.5를 만들기 위해 석회비료를 사용하는데 토양산도를 알고있을 때의 석회소요량을 연구결과 개발하였고 Eq. (15) 와 같다.

$$LR = \frac{A \times (X - Y) \times D}{10} \quad (15)$$

*LR* : Lime consumption

*A* : Amount of lime needed to increase pH 1.0.

*D* : Target acidity

*X* : Soil acidity

*Y* : Depth

Kim et al.(2018)은 토양수분 기반의 스마트 관개를 할 수 있는 자동 제어 시스템을 개발하는 연구를 하였다. 시스템은 크게 4가지, 데이터 수집부(토양수분-EC-지온센서, 감우센서), 데이터 송수신부(데이터로거, CDMA 모듈), 관개 시스템 제어부(컨트롤러, 전자밸브), 원격 모니터링 및 제어부(휴대폰 어플리케이션),로 구성되었다. 개발된 시스템을 오이 재배에

적용해 본 결과, 수확량이 21.3% 증가하였으며, 높은 품질의 상품이 관행 대비 33.7% 증가하였고 농가 소득이 25.1% 증가하였다고 보고하였다. 관계 소요 노동력이 월 19.5시간 절감하였으며, 농업용수 사용량이 약 11.0% 절약되었다.

노지스마트 팜에서 환경을 모니터링하고 활용한 사례를 정리하면 Table 1과 같다.

Table 1 Purpose, Network, measurement depth indicated in the reference

Thesis	Purpose	Network	Depth(cm)
Haggard(2010)	Providing information	WSN	10/30/50
Lee(2011)	Irrigation	USN	-
Chen(2014)	Irrigation	-	10/20/30/40
Youm(2018)	Irrigation / fertilizer application	Lora	-
Kim(2018)	Irrigation	CDMA	-
Goap(2018)	Irrigation	Zigbee	-

## 2.5. 노지에서 깊이별 지온데이터 예측 연구

Bilgili(2010)는 농업용 열펌프 시스템에서 지온의 예측이 중요하다는 사실을 인지하고 활용하기 위해 2000~2007년 동안 터키 Adana 기상관측소에서 측정한 깊이별 토양온도와 기상변수를 인공지능으로 학습시켜 5cm, 10cm, 20cm, 50cm, 100cm 깊이에서의 월별 토양온도를 예측하였다. 사용한 통계기법과 인공지능은 LR(Linear regression), NLR(Nonlinear regression), ANN(Artificial neural network) 세 가지 모델을 예측하는데 활용하였으며, 기온, 기압, 습도, 풍속, 강우량, 일사량, 일조시간을 매개변수로 활용하였다. 최적의 입력 변수를 선정하기 위해 추운 계절(10월~3월), 따뜻한 계절(4월~9월) 두시기로 나누어 각 기상인자들과 50cm 깊이에서의 토양온도 사이의 Cross-correlations를 계산하였다. 0.6이상의 값이 확인되는 변수를 적합한 변수라고 판단하였다. Bilgili(2010)의 실험결과에 따르면 동절기와 하절기별로 지온과 밀접한 상관을 갖는 변수는 약간의 차이가 있었는데, 그에 따르면 추운계절 에는 기온, 일조시간, 주기변수 월(month) 순서대로 0.6이상의 높은 상관관계가 있고 따뜻한 계절 에는 기온, 주기변수 월(month), 대기압, 일사량 순서대로 0.6이상의 높은 상관관계가 있었다. 강우량은 기준에 조금 못 미치는 -0.59의 상관계수를 나타내었다.

Bilgili(2010)의 연구에서는 기상인자 이외에 월(month)이라는 주기 변수를 활용하였고 이는 특정 깊이에서의 지온과 높은 상관관계가 있다고 보고되고 있다. 본 연구에서도 Bilgili(2010)의 연구를 참고하여 주기성을 갖는 인자들을 선정하여 활용하였고, 보다 구체적이고 연관성이 높은 주기성 인자를 매개변수로 활용하였다. 그뿐만 아니라 Bilgili(2010)는 터키 Adana 기상관측소에서 수집한 데이터를 Training과 Test로 나누어 예측 정확도를 확인하였는데, 본 연구에서는 실제 농업에서의 적용 가능성을 확인하기 위해 공공 기상대에서 관측한 자료를 통해 모델링하고 농업활동이 이루어지고 있는 밭 지역에서 검증을 진행하였다.

Alizamir(2019)는 관개를 효율적으로 하기 위해 기온, 습도, 일사량, 풍속 네 가지 기상인자와 인공지능 모델을 활용하여 5cm, 10cm, 50cm, 100cm 네 개의 깊이에서의 월별 토양온도 값을 예측하였다. 사용된 모델은 ELM(Extreme Learning Machine), ANN(Artificial Neural Networks), CART(Classification And Regression Trees), GMDH(Group Method of Data Handling) 네 가지 인공지능 모델이며 각 깊이와 기상인자 간의 산점도 그래프와  $R^2$ 값을 통해 상관관계 순서를 정하고 정해진 순서대로 하나씩 추가하여 정확도의 변화량을 확인하는 전진선택법(Feedforward Selection)을 통해 각 깊이에서 최적의 매개변수 조합을 선정하였다. 매개변수는 상관관계에 따라 조합을 통해 다음과 같이 매개변수 조합을 설정하였다.

1. 기온
2. 기온 + 습도
3. 기온 + 습도 + 일사량
4. 기온 + 습도 + 일사량 + 풍속

1986년부터 2010년까지 25년 동안의 월간 기록을 사용하였으며, 모델을 만드는데 사용한 training data와 test data의 비율은 8:2로 하였다. 정확도 평가는 RMSE, NSE, R2 Score 세 가지 수치로 평가하였다. 실험 결과 5cm, 10cm, 50 cm 깊이에서 단일 매개변수로 기온을 활용한 모델이 다른 모델들에 비해 우수한 정확도가 나오며, 100cm 깊이에서는 모든 인자를 매개변수로 활용한 모델이 가장 우수한 정확도가 나왔다고 보고하였다. 또한 토양의 깊이가 증가함에 따라 모델의 정확도가 감소한다고 보고하면서 50 cm 깊이까지는 기온 데이터만으로 성공적인 예측을 할 수 있으며 100cm 깊이에서 지온을 예측하려면 태양복사 및 풍속 정보가 필요하다고 주장하였다.

Alizamir(2019)의 연구에서 최적의 매개변수 조합을 찾기 위해 전진선택법을 적용하여 각 기상인자의 영향력의 크기와 여부를 확인하였다. 전진선택법의 경우 설명력에 따라 추가하는 변수의 순서가 정해지는데

Alizamir(2019)는 그 순서를 각 기상인자와 깊이별 토양온도간의 데이터를 단순 비교하여 산점도 그래프와  $R^2$ 값을 통해 설명력의 순서를 정하였다. 데이터 간 단순비교는 영향력이 높은 인자의 경우에는 단순비교를 통해 영향력을 확인하기 쉽지만 비교적 낮은 영향력을 갖는 인자의 경우에는 다른 인자의 방해받기 때문에 단순 비교만으로는 설명력을 확인하는 것이 어렵다. 따라서 본 연구에서는 Alizamir(2019)와 마찬가지로 전진선택법을 사용하였지만 각 기상인자를 하나씩 소거해가며 증감되는 정확도를 통해 각 인자의 설명력의 순서를 정하여 투입하였다.

Feng et al.(2019)은 2cm, 5cm, 10cm, 20cm 네 가지 깊이에서의 지온을 예측하기 위해 ELM(Extreme Learning Machine), GRNN(Generalized Regression Neural Networks), BPNN(Back Propagation Neural Networks), RF(Random Forests) 네 가지 머신러닝 모델의 예측 정확도를 확인하여 적용 가능성을 평가하였다. 기온, 상대습도, 일사량, 풍속, 증기압 다섯 가지 기상인자가 매개변수로써 활용되었으며, 중국 북부에 위치한 옥수수밭에서 2011년 1월 1일~2014년 12월 31일 4년 동안 연구가 수행되었다. 30분 주기의 데이터를 입력받아 학습시켰으며, training과 test를 75 : 25로 나누어 진행하였다. 정확도 평가는 Root mean square error (RMSE), mean absolute error(MAE), Nash - Sutcliffe coefficient (NS) and concordance correlation coefficient (cc) 네 가지 평가지표를 통해 확인하였다.

Feng et al.(2019)의 연구에서는 사용된 다섯 가지 기상인자에 대해서 예측에 미치는 영향력의 크기나 여부, 순서 등을 검증하지 않은 상태에서 예측을 진행하였다. 예측을 진행함에 있어 변수의 조합을 구성함에 따라 예측의 정확도가 달라지기 때문에 사용되는 매개변수가 어떻게 영향을 미치는지 확인해야 할 필요가 있다. 본 연구에서는 선정된 기상인자 및 주기성 인자의 영향력을 철저하게 검증하고 예측을 하였다.

## 2.6. 시계열 데이터 예측을 위한 연구

Zhou et al.(2019)는 중국 경제의 발전과 도시 확장으로 인해 심각해지고 있는 대기오염 문제를 해결하기 위한 일부의 역할을 위해 대기오염 농도 시계열 예측을 진행하였고, 이 결과를 분석하였다. 머신러닝 모델 GRU를 사용하여 봄·여름·가을·겨울 사계절에 따라 4개 모델을 교육하고, 계절 별 오후 2시 30분에서의 예측에 대한 4개 모델의 효과를 Test data와 비교하여 평가하였다. 논문에 사용된 매개변수 데이터 세트는 두 부분으로 구성된다. 2010년부터 2014년까지 베이징에서 오후 2시 30분 시간대에서의 시간당 평균 대기 오염 농도와 2010년부터 2014년까지의 베이징 기상 데이터이다. 기상인자로서는 이슬점, 기온, 기압, 풍속, 강설량, 강수량, 풍향이 사용되었다. 또한 데이터세트는 시간대에 따라 나누어 진행하였는데, 봄(3월, 4월, 5월), 여름(6월, 7월, 8월), 가을(9월, 10월, 11월), 겨울(12월, 1월, 2월)으로 나누었다. 모델의 정확성을 평가하기 위해서는 MAE, MSE, RMSE를 사용하였고, 실험 비교 분석 결과 MAE의 효과가 다른 두 가지에 비해 우수하다고 판단하였다.

Zhou et al.(2019)의 경우 시계열 데이터에 영향을 미치는 계절성을 적용하기 위해 1년을 4분기로 데이터 세트를 나누어 따로 실험을 진행하였다. 네 개로 분리하여 실험을 진행하게 되면 실험 시간의 소모가 네 배로 증가하게 되어 각 매개변수의 영향력을 평가하기 위해 여러 개의 매개변수 조합을 실험하는 본 연구에 적용하기에는 그만큼 시간이 많이 소요된다. 본 실험에서는 데이터 세트를 계절별로 분리하여 진행하지 않고 계절성을 나타낼 수 있는 매개변수를 투입하여 실험을 진행하였다.

Lee et al.(2021)는 지구온난화로 인해 발생된 기온 상승에 따른 극한 기후에 대한 대책을 사전에 마련하기 위해 연평균 기온 상승률과 변화량이 높은 지역에서의 정확도 높은 평균기온 예측 연구가 이루어져야 할 필요성을 파악했고, 시간의 흐름에 따라 변화하는 데이터를 처리하는 데 좋은 성과를 보이는 RNN(Recurrent neural network) 모델과 RNN이 가지

고 있는 장기의존성(Long-Term Dependency)문제를 해결하기 위해 고안된 LSTM(Long Short-Term Memory) 모델과 GRU(Gated Recurrent Unit) 모델을 통해 연평균기온 상승률이 높은 강원도 원주시를 대상으로 평균기온을 예측하고 정확도를 평가하는 연구를 진행하였다. 각 실험모델의 평가를 위한 loss값은 MSE(Mean Squared Error)로 진행하였고, 평가는 MAE((Mean Absolute Error)를 활용하였다. Epoch가 100으로 설정되어 깊은 학습이 진행되지 못하여 모든 분석에서 전체적으로 높은 정확도를 확인하지 못하였고, LSTM 모델과 GRU 모델의 결과값의 차이가 거의 없었다. 실험을 통해 충분한 학습 횟수와 변수의 간소화가 예측 정확성에 미치는 영향력을 확인할 수 있었다. 모델의 성능을 테스트하는 데 사용되는 데이터 세트는 ISRO(Indian Space Research Organisation)의 기록을 활용하였다. 테스트 결과 각 모델의 Loss값은 RNN 1.1316, LSTM 0.8638, GRU 0.9183으로 출력되어 LSTM은 고려된 세 가지 모델 중 검증 손실이 가장 적으며 SimpleRNN이 그 뒤를 이으며 검증 손실이 가장 높은 모델은 GRU 임을 확인 할 수 있었고, 모델의 정확도를 확인할 수 있는 MSE의 경우 RNN 8.178, RNN 16.6962, GRU 18.7785로 LSTM이 두 모델보다 높은 정확도로 예측이 가능하다는 것을 확인할 수 있었다. 모델의 시간소비를 고려하였을 때 RNN이 시간이 가장 많이 걸렸으며, GRU와 LSTM은 실행시간이 거의 동일하였지만 게이트 구조가 단순한 GRU가 LSTM보다 일반적으로 빠른 것으로 확인 할 수 있었다.

본 연구에서는 Lee et al.(2021)가 겪었던 오류를 범하지 않기 위해 학습 시 Epoch를 300으로 설정하였고, 그에 따른 Overfitting의 발생을 방지하기 위해 early stopping 명령을 활용하여 validation loss가 증가하는 시점에서 훈련을 멈추도록 설정하였다.

### 3. 재료 및 방법

본 연구에서는 깊이별 지온을 예측하기 위하여 적절한 기상인자와 시계열 변수를 선정하고, 선택된 세 가지 인공지능과 공공 기상자료를 이용하여 학습한 뒤에 실측된 지온 데이터를 이용하여 최적의 매개변수와 모델 조합을 평가하고 선택된 최적의 조합과 실측값을 비교하여 적용 가능성을 평가하는 방법을 이용하였다.

하나의 데이터 세트를 이용해 학습 데이터와 평가 데이터로 일정 비율 나누어 정확성을 평가하였던 기존의 연구들과 다르게 별개의 평가 데이터를 사용하였으며, 그 이유는 시간과 비용을 들이지 않고 농업현장에서 사용하려면 무료로 제공되는 공공데이터를 이용할 수밖에 없기 때문이다.

뿐만 아니라 본 연구에서는 주기성을 갖는 인자가 투입되어 예측 정확도에 미치는 영향을 확인하였다. 토양온도와 마찬가지로 연간 주기나 주행성을 갖는 인자를 매개변수로 투입하여 예측의 정확도를 높이는데 미치는 영향의 여부를 확인하고자 하였다. 깊이별 토양온도를 예측하였던 기존의 연구들과 본 연구의 데이터세트 활용의 차이는 Figure 4 를 통해 확인 가능하다.



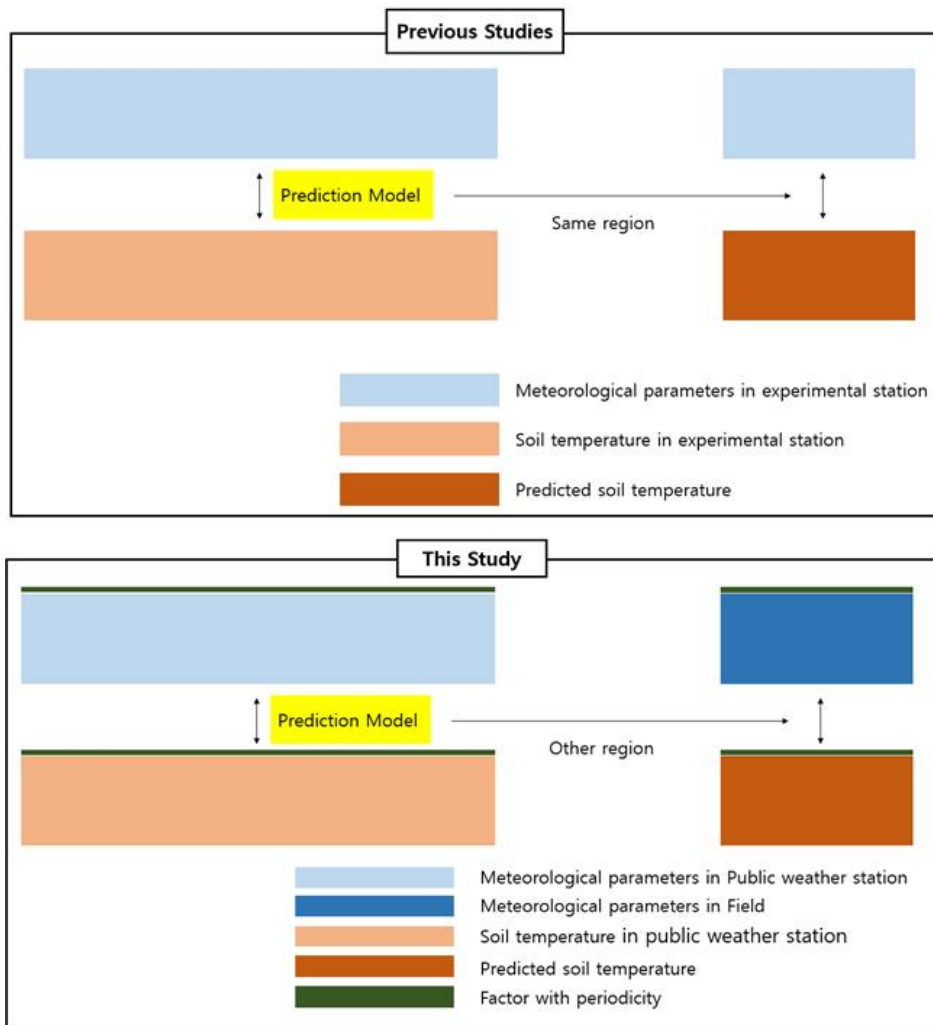


Figure 4 The experimental methods of previous experiments and the experimental methods of this experiment

### 3.1. 공공기상자료 확보

#### 3.1.1. 공공기상대 데이터

예측 모델을 만드는데 활용한 데이터는 2016년 1월부터 2020년 12월까지 5년간 춘천 기상대, 인천 기상대, 서울 기상대, 청송 기상대, 대전 기상대, 고창 기상대, 순창 기상대, 의령 기상대, 보성 기상대, 김해 기상대 10개 지역에서의 기상 및 지온 데이터는 ‘기상청’에서 운영하는 ‘기상자료 개방 포털’ 사이트에서 취득하였으며, 각 기상대의 위치는 Figure 5 와 같다.

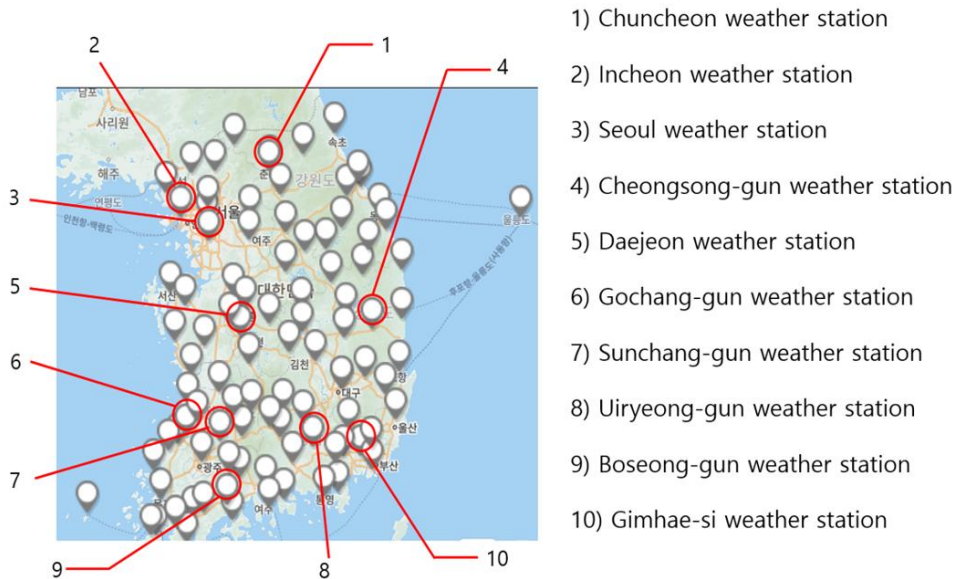


Figure 5 Data collected from public weather stations.

우리나라에는 2021년 현재 전국에 102개의 종관 기상관측 공공 기상대가 있고 지온 정보를 제공하는 기상대는 그중 28개소가 있기 때문에 공공 기상대가 측정시설을 지속적으로 확충한다면 본 연구결과는 전국 농경지에 적용될 수 있음을 의미한다. 본 연구에서 선정한 공공 기상대의 지리 정보와 기상특성은 Table 2 와 같다.

Table 2 latitude, longitude, altitude above sea level, and climatic characteristics of each public weather station.

Weather station	Latitude/Longitude	Elevation	Local climate
Chuncheon weather station	37.9026 / 127.7357	76m	Inland
Incheon weather station	37.47772 / 126.6249	69m	Continental
Seoul weather station	37.57142 / 126.9658	86m	Continental
Cheongsong-gun weather station	36.4351 / 129.04005	209m	Continental
Daejeon weather station	36.37198 / 127.37211	70m	Continental
Gochang-gun weather station	35.42668 / 126.69724	59m	Continental
Sunchang-gun weather station	35.37131 / 127.1286	129m	Inland
Uiryeong-gun weather station	35.32258 / 128.28812	14m	Oceanic
Bosung-gun weather station	34.76335 / 127.21226	1m	Oceanic
Gimhae weather station	35.22981 / 128.89075	55m	Oceanic

### 3.1.2. 학습용 데이터의 전처리

5년간 10개소의 1시간 간격으로 총 438,481개의 데이터를 이용하였으나 기상대의 사정에 따라 데이터 결측 시간대가 존재하며 이런 자료가 인공지능 학습에 오차를 유발할 수 있으므로 결측된 시간대는 제외하며 실험하였다. 전처리를 통해 제외된 데이터의 개수와 크기는 매개변수 조합에 따라 다르다.

## 3.2. 지온예측 평가용 지온데이터

공공기상 데이터와 인공지능 모델의 성능을 평가하기 위해 모델의 input으로 시험포장의 기상 데이터가 사용되었으며, 예측된 깊이별 지온값의 정확도를 확인하기 위해 사용된 실측값으로는 시험포장의 깊이별 지온 데이터가 사용되었다.

### 3.2.1. 외부기상대 및 토양센서

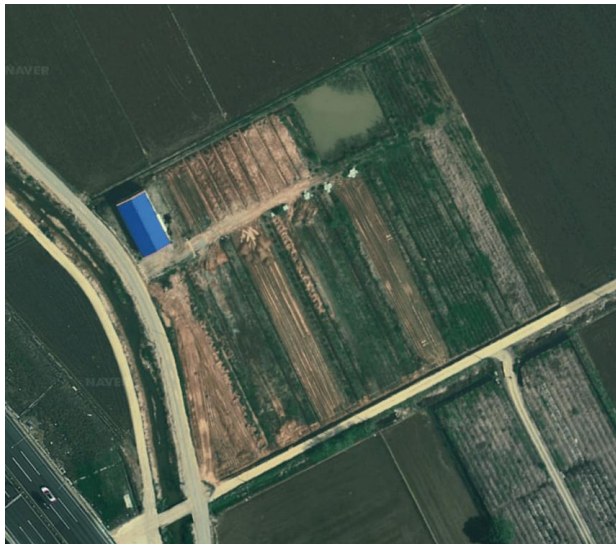


Figure 6 Satellite Image of Test Field  
(Source : Naver Map)

평가 데이터로 활용될 시험포장이 위치하는 곳의 주소는 아산호 주변 객토한 밭으로 매년 감자, 대파, 콩 등 노지작물의 재배가 이뤄지고 있는 밭이다. 아산은 차령산맥의 북쪽에 차령산맥과 인접하여 자리하지만 서해안에 가까이 있어서 내륙 지역에 비해 1월 기온이  $-2.1^{\circ}\text{C}$ 로 온대의 경계인  $-3^{\circ}\text{C}$ 를 약간 넘기 때문에 평균기온으로 보면 온대기후에 해당하지만 바다의 영향이 지역에 따라 차이가 나서  $-3^{\circ}\text{C}$  등온선이 삼교천과 평행하

계 남북으로 지나는 특성이 있다. 연평균 강수량은 1,365.1mm로 충청남도 지역 평균보다 약간 많은 편인데 이는 바다와 인접하였고 지형이 서해안 쪽으로 뚫려 있어서 습기 유입이 많으며, 서해안 일대에서는 비교적 높은 산지인 광덕산, 봉수산 등이 있기 때문이다.

인공지능 모델과 공공 기상 데이터를 통해 만들어진 예측 모델로 실험이 진행되는 포장에서 깊이별 지온을 예측하는데 필요한 포장의 기상 데이터를 취득하기 위한 노지용 외부 기상대는 스마트팜 기업 ‘나래트랜드’사에서 제공하는 ‘과수노지 ICT 생육환경 모니터링 시스템’을 사용하였으며, 제품에 포함되어 있는 각 센서의 사양은 Table 3, 외부 기상대와 토양 센서의 사진은 Figure 7 과 같다.

Table 3 Specifications of each weather and soil sensor.

Sensor	range	Precision
Air temperature	-25~65°C	±0.1°C
Humidity	0~100%	±3%
Wind direction	0 ~ 360 °	±7°
Wind speed	0-50 m/s	1m/s
insolation	0~2,000 W/m <sup>2</sup> (±5%)	±5%
Soil	-10°C to +55°C	±0.3°C



Figure 7 Picture of outside expectation (left) and soil sensor (right)

검증용 밭 내의 외부 기상대 및 토양 센서가 설치된 위치와 그 모식도는 Figure 8 와 같다.

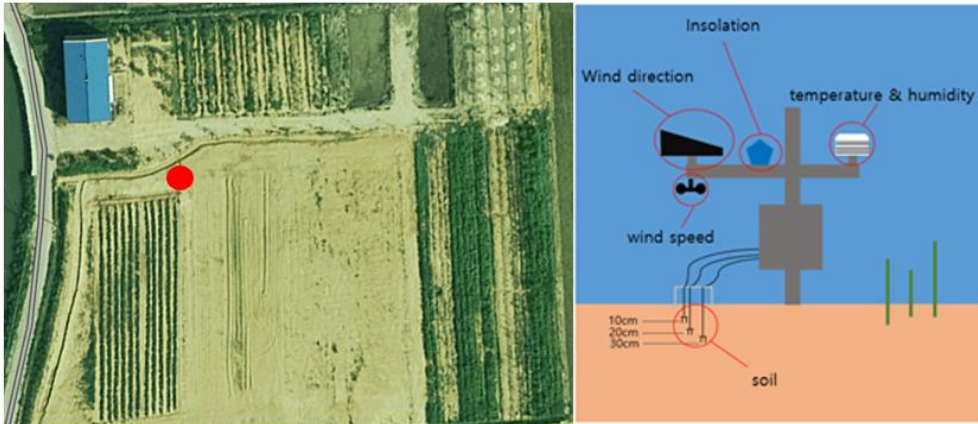


Figure 8 Location of external expectations and soil sensors(left) and schematic diagram(right) in the field

### 3.3. 매개변수 선정

본 연구에서 토양 깊이별 온도를 예측하는 데 사용한 매개변수는 기상과 관련된 인자와 주기성을 대표하는 매개변수로 구성하였다. 기상과 관련된 인자는 섭씨로 표시한 기온과 % 단위로 표시한 상대습도, m/s 단위로 표시한 풍속, degree로 표시한 풍향, MJ/m<sup>2</sup> 단위로 표시한 일사량이었다. 주기성을 나타내는 매개변수에는 전국 섭씨로 표시한 월평균기온, radian 단위의 태양고도, 섭씨로 표시한 타지역의 깊이별 지온 데이터를 활용하였다.

Table 4 The parameters used in the experiment

parameter	Unit	Classification
Air temperature	℃	Weather
Humidity	%	Weather
Wind direction	°	Weather
Wind speed	m/s	Weather
Insolation	MJ/m <sup>2</sup>	Weather
Average monthly temperature nationwide	℃	Periodicity
Altitude of the sun	radian	Periodicity
Soil temperature in other region	℃	Periodicity

깊이별 지온의 추세를 확인하기 위한 매개변수로서 타지역의 깊이별 지온 데이터를 활용하였다. 지온은 기후의 영향을 받으며, 전국 각 지역의 기후는 차이가 있지만 비슷한 추세를 보이고 있다. 마찬가지로 공공기상 데이터의 자료에 의하면 전국 깊이별 지온 변화량은 차이가 있지만 비슷한 추세를 보이고 있기 때문에 모델을 만들거나 예측과 비교한 실측 지온 데이터 이외의 깊이별 지온 데이터를 매개변수로서 활용하였다. 데이터는 기상청에서 운영하는 기상자료개방포털(data.kma.go.kr)에서 제공하는 서산 기상대 공공데이터의 10cm, 20cm, 30cm의 토양온도 데이터를 사용하였다.

### 3.4. 인공지능 모델 선정과 학습

#### 3.4.1 RNN(Recurrent neural network)

RNN(Recurrent neural network) 모델은 은닉층의 노드에서 활성화 함수를 통해 나온 결과값을 출력층 방향으로 보내면서, 다시 은닉층 노드의 다음 계산의 입력으로 보내는 특징을 갖고 있는 시퀀스 모델이다. 몇 단계 전으로 거슬러 올라가 데이터를 반영하는 신경망 알고리즘이기 때문에 1회의 데이터가 아닌 과거의 데이터도 학습에 활용이 가능하다. RNN은 시계열 데이터 학습이 가능하며, 음성 파형이나 자연어 학습이 가능하기 때문에 변수의 시간적 변동에 따라 값의 예측이 필요한 본 연구에 적합한 모델이다.

다만 RNN은 현재 노드와 먼 과거 상태를 사용한 문맥의 처리가 어렵다. 이러한 문제를 장기 의존성 문제(Long-Term Dependency)라고 하는데 이러한 문제를 보완하기 위해 LSTM(Long Short-Term Memory) 모델이 개발되었다.

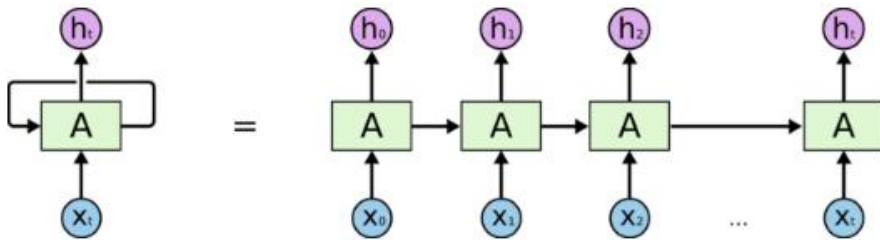


Figure 9 RNN Model Structure(Colah's Blog, 2015)



### 3.4.2 LSTM(Long Short-Term Memory)

RNN 모델의 경우 각 시간대에 동일한 작동을 반복적으로 적용해야 되기 때문에 시점이 길어질수록 앞의 정보가 뒤로 충분히 전달되지 못하여 시간대가 길어질수록 정보가 충분히 전달되지 못해 정확도가 떨어지는 장기의존성(Problem of Long-Term Dependencies) 문제가 발생하게 된다(, 2019). LSTM(Long Short-Term Memory) 모델은 이러한 장기의존성 문제를 해결하기 위해 개발된 모델이다.

LSTM의 각 메모리 셀은 잊기 게이트, 입력 게이트, 출력 게이트라고 불리는 세 개의 "게이트"에 의해 업데이트되도록 제어된다. 각 게이트의 출력은 시그모이드 함수에 의해 계산된 0과 1 사이의 숫자이며, 이 숫자는 데이터가 각 게이트를 통과할 수 있는 범위를 나타낸다. LSTM 유닛은 정보가 저장된 메모리 셀에 의해 구성되며, 이러한 LSTM은 긴 시간 지속성을 가지고 있다. LSTM 네트워크는 복잡한 다변량 시퀀스에서 장기 시간 의존성을 캡처할 수 있기 때문에 시계열 모델링에 매우 유용하다 (Siarni-Namini, 2018).

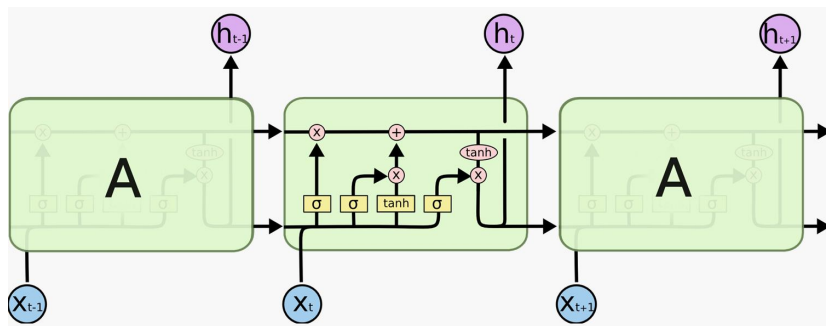


Figure 10 LSTM Model Structure(Colah's Blog, 2015)

### 3.4.3 GRU(Gated recurrent unit)

GRU(Gated recurrent unit) 모델은 LSTM 모델의 장기 의존성 문제에 대한 해결책을 유지하여 LSTM과 유사한 성능을 유지하며 구조를 단순하게 만드는 LSTM 모델의 변형 형태이다(Zhou et al., 2019). LSTM에서는 출력, 입력, 삭제 게이트라는 3개의 게이트가 존재한 반면, GRU에서는 업데이트 게이트와 리셋 게이트 두 가지 게이트만이 존재하기 때문에 일반적인 경우에서 GRU는 LSTM보다 학습 속도가 빠르고 비슷한 성능을 보인다.

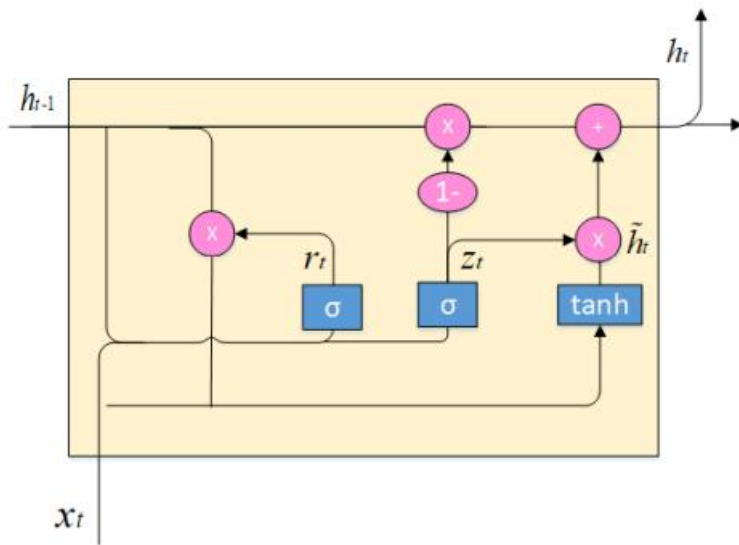


Figure 11 GRU Model Structure(Zhou, 2019)

선정된 인공지능 알고리즘을 이용한 예측모형은 프로그래밍 언어인 Python을 사용하여 개발되었다. 각 예측모형은 1개의 은닉층(Hidden layer)을 구성하여 만들었으며, 데이터 세트는 Training, Validation, Test 으로 6:2:2의 비율로 나누어 진행하였다.

### 3.5. 깊이별 지온 예측 절차 및 평가 방법

실험은 1) 기상인자 데이터를 이용한 최적의 매개변수 선택, 2) 주기성을 나타내는 매개변수 선택, 3) 최적의 매개변수를 이용한 예측값과 실측값 비교, 크게 세 가지로 구분하여 진행 하였다.

실험 1)의 경우 각 기상인자 매개변수를 하나씩 소거하여 예측한 예측값의 정확도를 다섯 가지 기상인자를 전부 활용하여 예측한 예측값의 정확도와 비교하여 각 기상인자가 깊이별 지온을 예측 하는데 미치는 영향력의 크기와 순서를 확인하였고, 전진선택법(Feedforward Selection)을 통해 영향력의 순서에 따라 매개변수로 투입하여 정확도의 증감을 확인하며 최적의 기상인자 매개변수 조합을 선택 하였다.

실험 2)에서는 기상인자 이외에 주기에 따라 패턴을 가지고 있는 주기성 매개변수 세 가지를 각각 실험 1)에서 선택된 기상인자 매개변수 조합에 투입시켜 예측값의 정확도를 증가시키는데 기여하는지 여부를 확인하고 모든 깊이에서 가장 성능이 좋다고 평가되는 기상인자 매개변수, 주기성 매개변수, 인공지능 모델 조합 두 개를 선택 하였다.

실험 3)에서는 실험 2)를 통해 선택된 최고의 성능을 가진 두 개의 조합을 실측값의 패턴분석, 잔차분석 을 통해 유사성을 평가하고, 농업에서의 활용 가능성 여부를 평가하였다.

인공지능 알고리즘별 예측치에 대한 평가는 정확도는 Root Mean Square Error(RMSE), Nash - Sutcliffe Efficiency(NSE), Determination Coefficient( $R^2$ ) 세 가지 지표의 값을 통해 평가하였다.

$$\text{Root Mean Square Error(SE)} = \sqrt{\frac{\sum_{i=1}^N (ST_{im} - ST_{ip})^2}{N}} \quad (16)$$

$$\text{Nash - Sutcliffe Efficiency(NSE)} = 1 - \frac{\sum_{i=1}^N (ST_{im} - ST_{ip})^2}{\sum_{i=1}^N (ST_{im} - \overline{ST_m})^2} \quad (17)$$

$$\text{Determination Coefficient}(R^2) = \frac{\sum_{i=1}^N (ST_{im} - \overline{ST_m})(ST_{ip} - \overline{ST_p})^2}{\sum_{i=1}^N (ST_{im} - \overline{ST_m})^2 \sum_{i=1}^N (ST_{ip} - \overline{ST_p})^2} \quad (18)$$

## 4. 결과

### 4.1. 기상인자 데이터를 이용한 최적의 매개변수 선택

#### 4.1.1. 소거를 이용한 기상인자 영향력 순서 평가

Table 5 Combination of parameters of each dataset  
for experiment 4.1.1

Data set	Combination of parameters
Set_1	Air temperature / Humidity / Wind direction / Wind speed / Insolation
Set_2	Humidity / Wind direction / Wind speed / Insolation
Set_3	Air temperature / Wind direction / Wind speed / Insolation
Set_4	Air temperature / Humidity / Wind speed / Insolation
Set_5	Air temperature / Humidity / Wind direction / Insolation
Set_6	Air temperature / Humidity / Wind direction / Wind speed

각 기상인자가 매개변수로써 활용되었을 때 예측 정확도에 미치는 영향력을 확인하기 위해 여섯 개의 매개변수 조합을 Table 5 와 같이 구성하였다. 선정된 모든 기상인자들의 조합으로 데이터세트를 구성한 Set\_1과 다섯 개의 기상인자를 하나씩 소거하여 구성된 그 외의 조합으로 구성하였다. 각 조합을 이용해 깊이별, 모델별로 예측 모델을 만들어 각 예측값과 실측값을 비교하였다. 예측의 정확도를 높이는데 큰 영향력을 미치는 기상인자일수록 제외했을 시 산점도 그래프와 각종 수치에서 정확도의 감소를 나타낼 것이고, 적은 영향력을 미치는 기상인자라면 그 반대의 결과를 나타낼 것이다. 각 깊이별, 데이터 세트 별 실측값-예측값 산점도 그래프와 관계식 및  $R^2$  값은 다음과 같다.

10cm - RNN

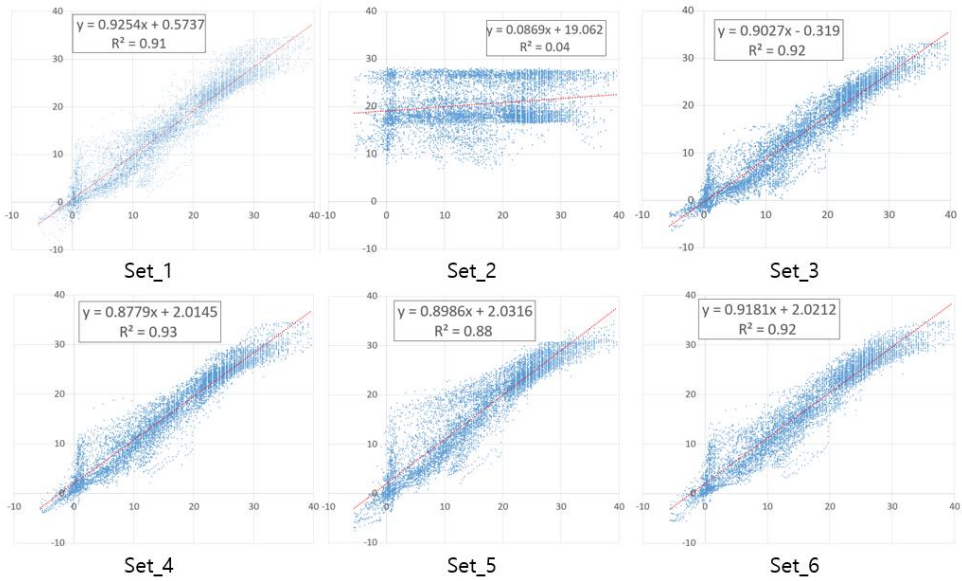


Figure 12 Scatter plot graph of each set of RNN models at a depth of 10 cm

10cm - LSTM

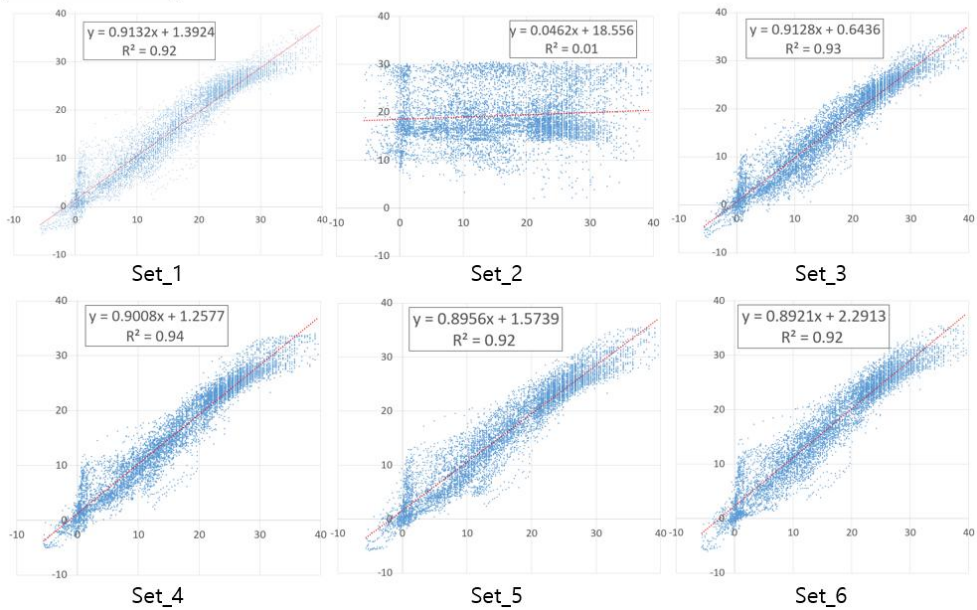


Figure 13 Scatter plot graph of each set of LSTM models at a depth of 10 cm

10cm - GRU

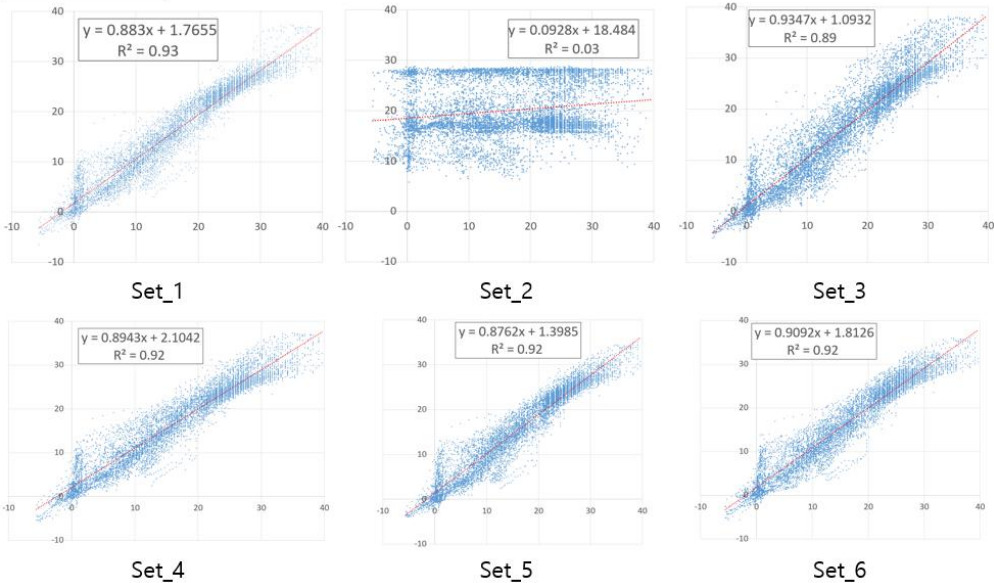


Figure 14 Scatter plot graph of each set of GRU models at a depth of 10 cm

10cm 깊이에서 실측값-예측값 산점도 그래프를 확인한 결과 Set\_2를 제외한 모든 데이터세트의 산점도 그래프는 모든 모델에서 선형성을 보이며  $R^2$  값이 0.85 이상으로 높은 정확도를 보였다. Set\_2의 산점도 그래프는 모든 모델에서 그래프의 성형성을 보이지 못하고  $R^2$  값이 0.1이하로 낮은 정확도를 보이는 것을 확인하여 Set\_2의 예측이 제대로 진행되지 않았음을 확인할 수 있다.

10cm 깊이에서의 예측은 기온이 가장 영향력을 크게 미치는 것으로 그래프를 통해 평가할 수 있다. 기온 이외의 기상인자의 영향력 크기의 순서는 평가 지표의 수치 비교를 통해 평가해야 할 필요가 있다.

20cm - RNN

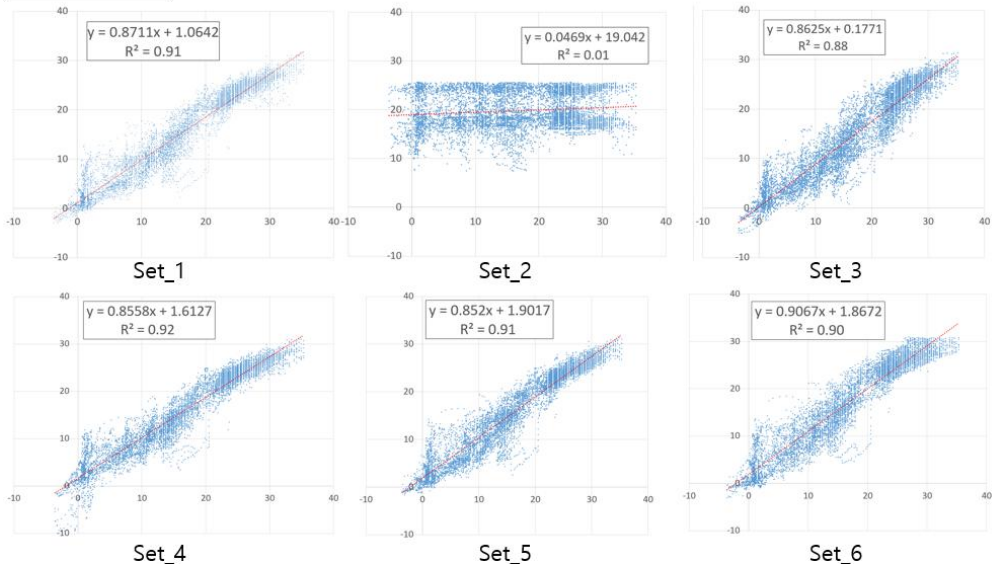


Figure 15 Scatterplot graph of each set of RNN models at a depth of 20 cm

20cm - LSTM

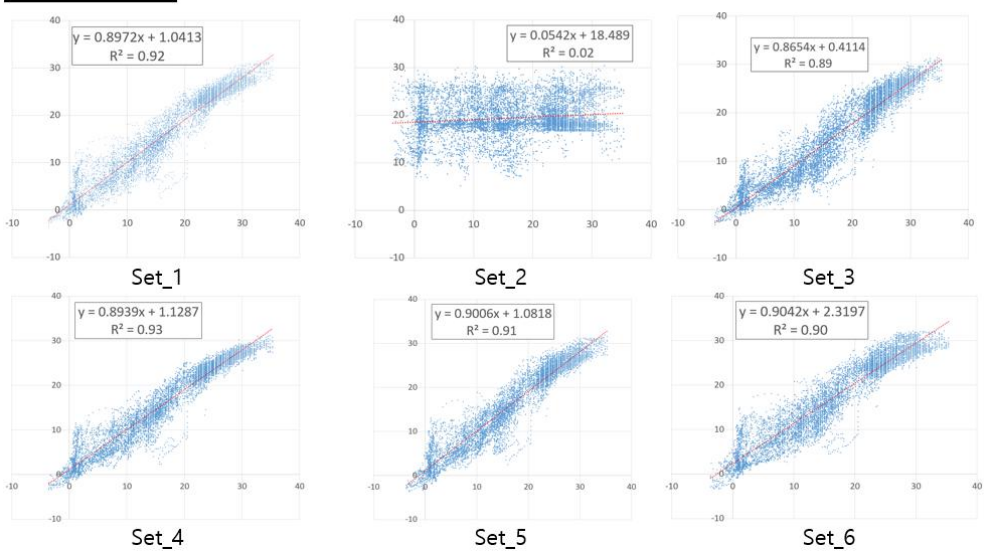


Figure 16 Scatterplot graph of each set of LSTM models at a depth of 20 cm

20cm - GRU

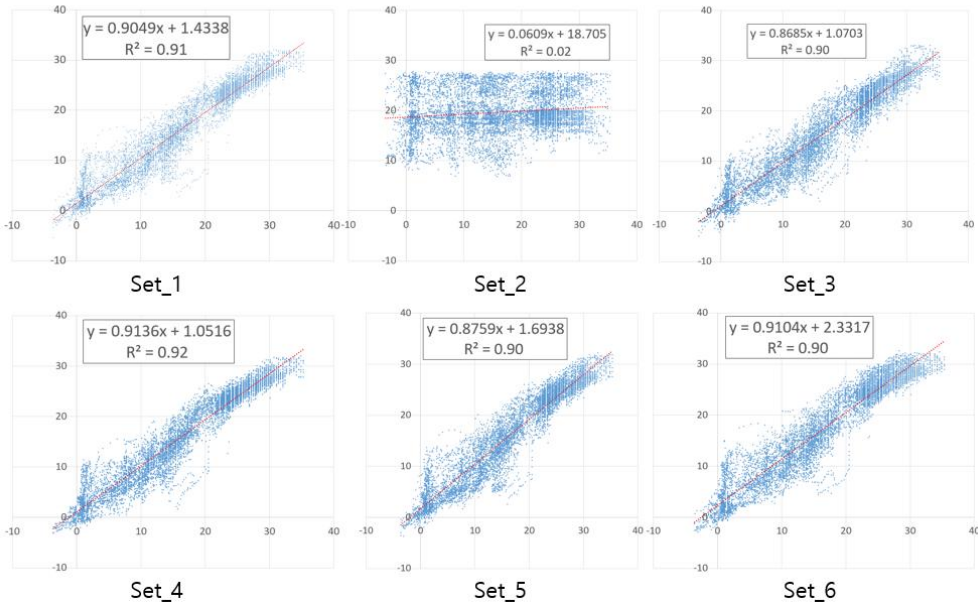


Figure 17 Scatterplot graph of each set of GRU models at a depth of 20 cm

20cm 깊이에서도 같은 양상을 보였는데, 10cm 깊이에서의 예측과 마찬가지로 Set\_2를 제외한 모든 데이터세트의 산점도 그래프는 모든 모델에서 선형성을 보이며  $R^2$ 값이 0.85 이상으로 높은 정확도를 보이는 것을 확인하였다. Set\_2 역시 10cm에서의 예측과 마찬가지로 모든 모델에서의 그래프는 선형성을 보이지 못하고  $R^2$ 값이 0.1이하로 낮은 정확도를 보이는 것을 확인하였다.

다만 20cm의  $R^2$ 값은 평균적으로 소폭 감소한 것을 확인할 수 있는데, 이는 깊이에 증가함에 따라 지표면에서 교환된 열 교환의 영향력이 토양의 비열로 인해 감소되었음을 확인할 수 있다.

20cm 깊이에서의 예측에서도 기온이 가장 영향력을 크게 미친다고 평가할 수 있으며, 기온 이외의 기상인자의 영향력 크기의 순서는 평가 지표의 수치 비교를 통해 확인해야 할 필요가 있다.



30cm - RNN

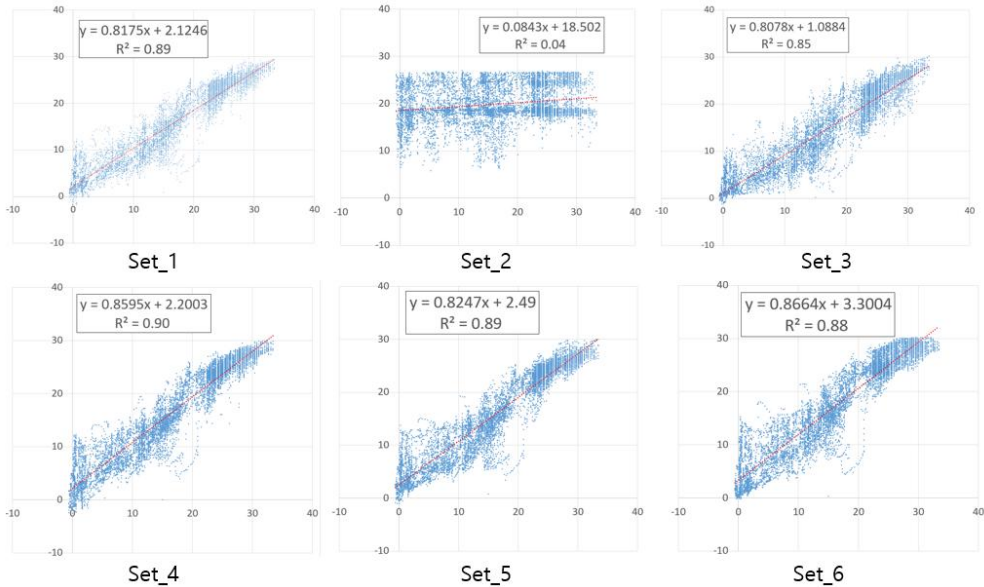


Figure 18 Scatterplot graph of each set of RNN models at a depth of 30 cm

30cm - LSTM

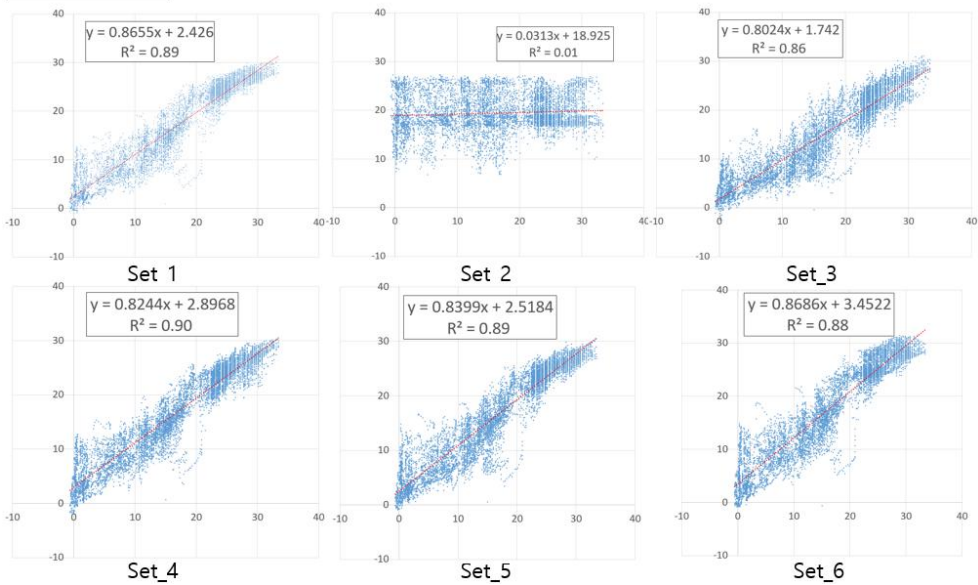


Figure 19 Scatterplot graph of each set of LSTM models at a depth of 30 cm

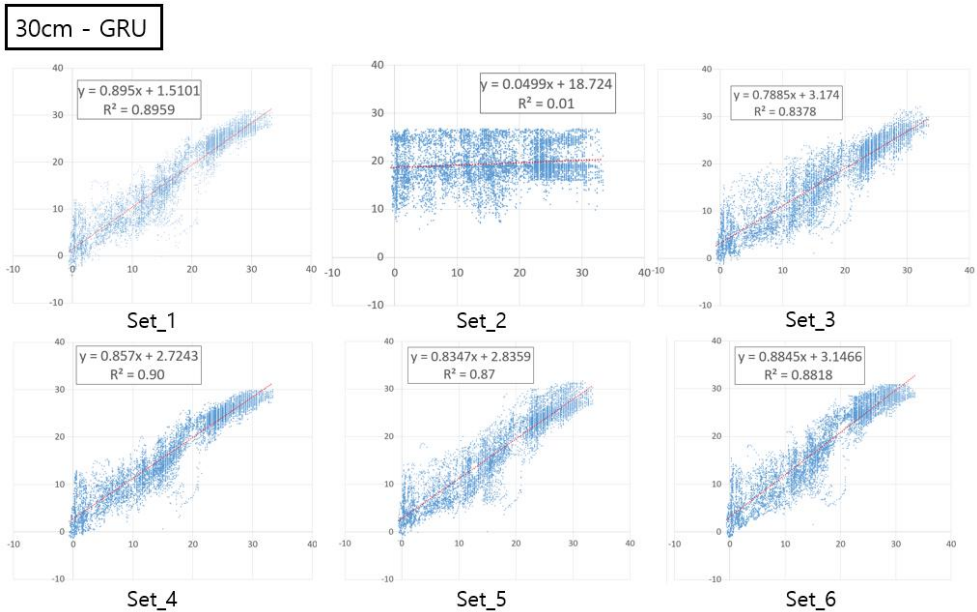


Figure 20 Scatterplot graph of each set of GRU models at a depth of 30 cm

30cm 깊이에서도 10cm, 20cm 깊이에서의 예측과 마찬가지로 Set\_2를 제외한 모든 데이터세트의 산점도 그래프는 모든 모델에서 선형성을 보이며  $R^2$ 값이 0.85 이상으로 높은 정확도를 보이는 것을 확인하였다. Set\_2 역시 10cm에서의 예측과 마찬가지로 모든 모델에서 선형그래프는 보이지 못하고  $R^2$ 값이 0.1이하로 낮은 정확도를 보이는 것을 확인하였다.

30cm 깊이에서의 예측 역시 기온이 가장 영향력을 크게 미친다고 평가할 수 있으며, 기온 이외의 기상인자의 영향력 크기의 순서는 평가 지표의 수치 비교를 통해 확인해야 할 필요가 있다는 것을 확인하였다.

세 깊이에서의 실측값-예측값 산점도 그래프를 통해 확인할 수 있었던 것은 기온이 깊이별 지온을 예측하는 데 있어 가장 큰 영향력을 미친다는 것이다.

기온 이외의 기상인자의 영향력의 크기와 순서는 다른 깊이와 모델에서 각 세트별 평가 지표의 수치를 비교하여 평가하였다. 각 평가 지표를 비교한 그래프는 다음과 같다.

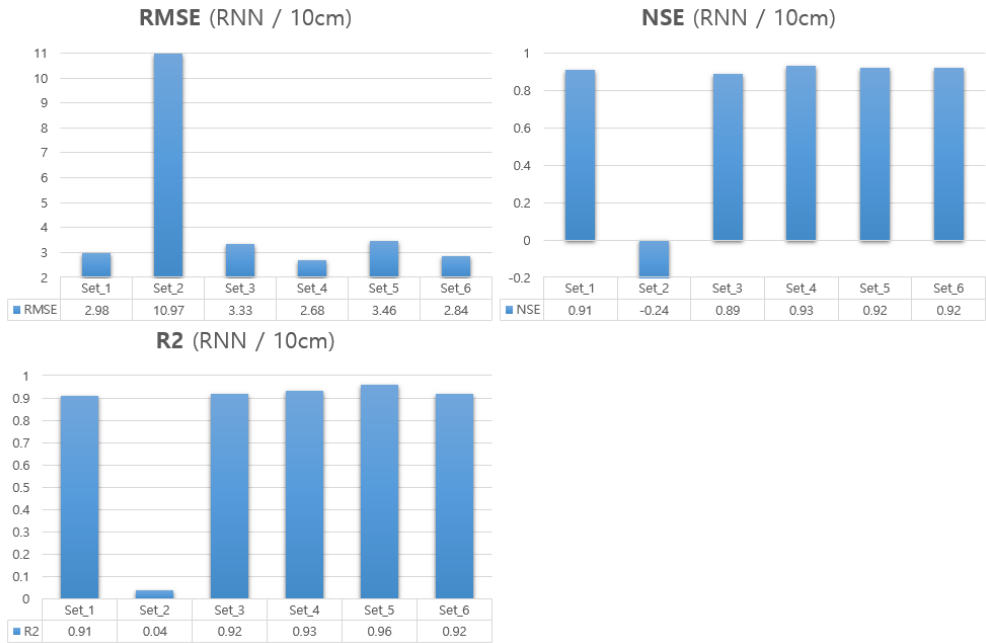


Figure 21 RMSE, NSE, and R2 values of RNN model predictions at a depth of 10 cm.

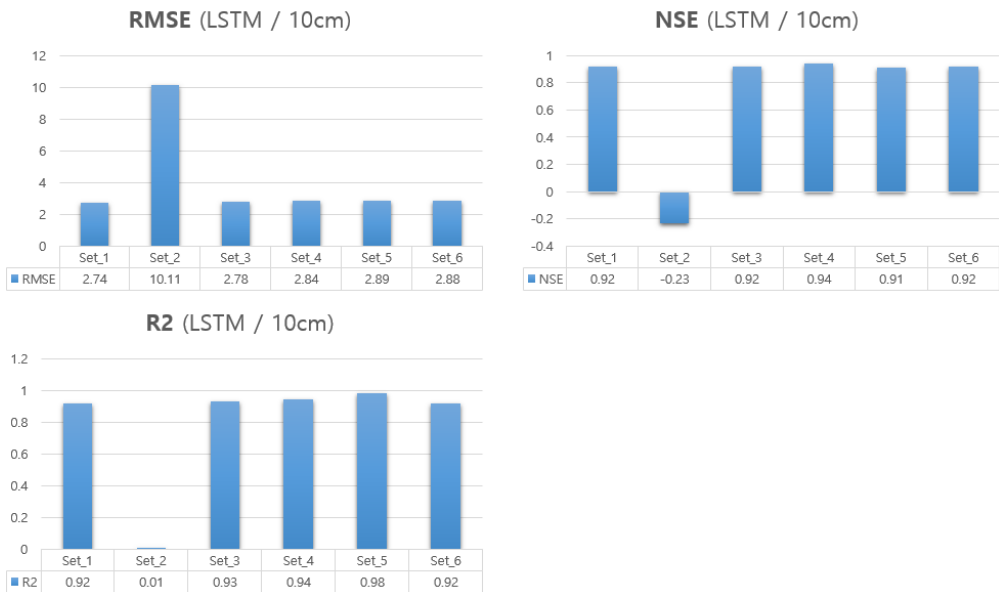


Figure 22 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 10 cm.

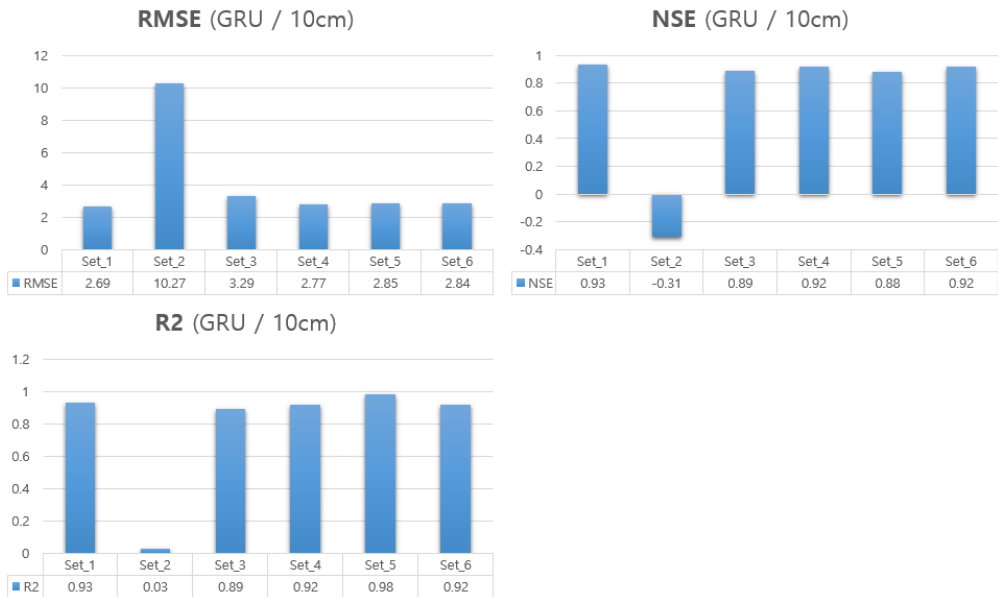


Figure 23 RMSE, NSE, and R2 values of GRU model predictions at a depth of 10 cm.

10cm 깊이에서 각 모델 별 예측값을 평가 지표를 통해 평가하였을 시 Set\_2의 정확도가 큰 차이로 가장 낮고, Set\_1과 Set\_4의 정확도가 가장 높은 것을 확인할 수 있었다. Set\_3, Set\_5, Set\_6의 경우에는 사용 모델과 평가 지표에 따라 정확도의 순서의 차이가 있었다.

실험 결과를 통해 판단할 수 있는 사실은 10cm 깊이에서 인공지능을 활용하여 지온 데이터를 예측할 때 기온이 가장 큰 영향력을 미치며, 풍향은 가장 영향력이 작거나 영향력이 없다는 것이다.

10cm 깊이에서 습도, 풍속, 일사량의 경우 영향력은 미치지만 비교적 크지 않으며, 세 가지 기상인자의 영향력의 우위는 정의하기 어렵다.

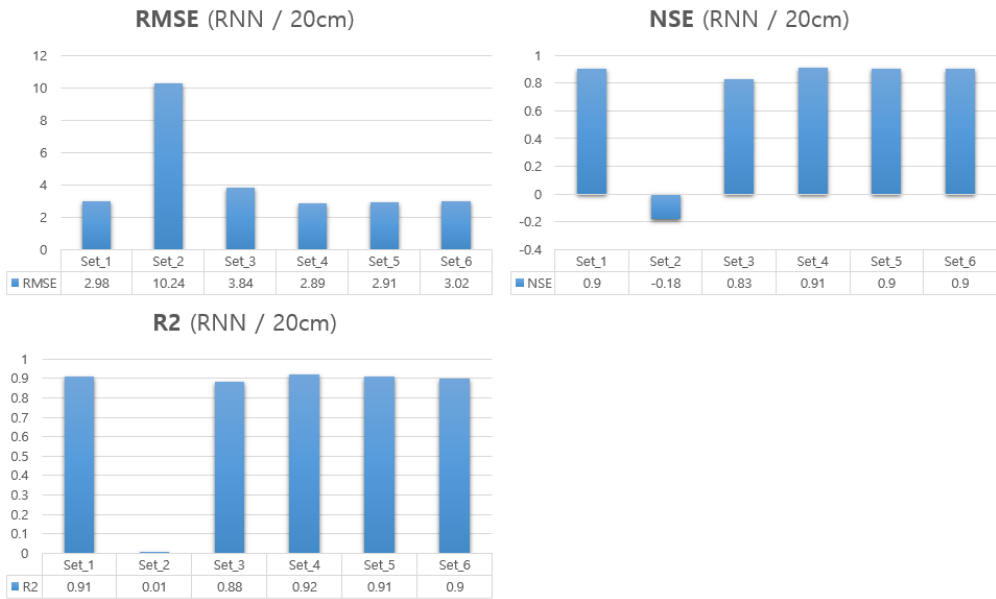


Figure 24 RMSE, NSE, and R2 values of RNN model predictions at a depth of 20 cm.

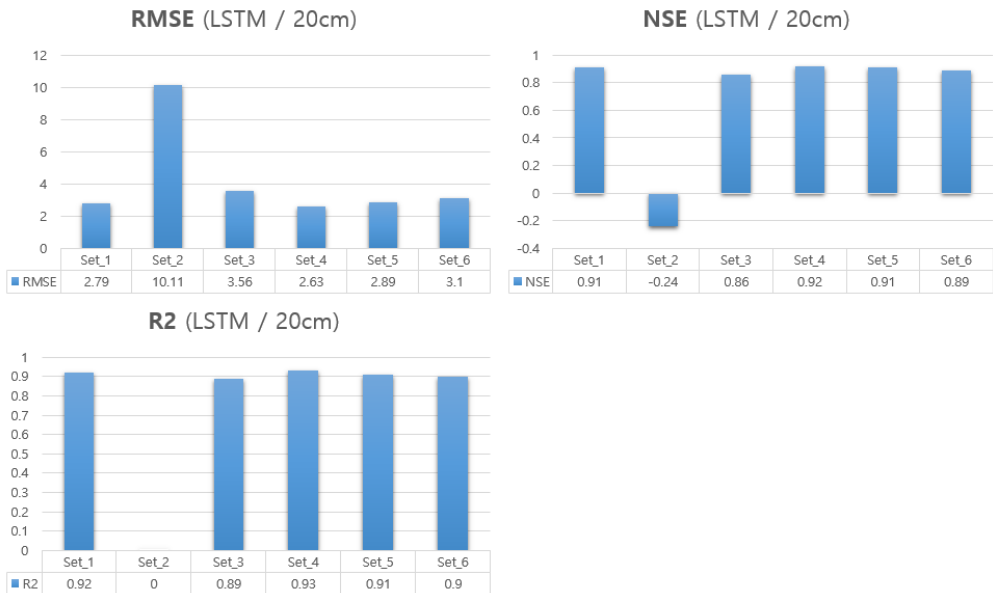


Figure 25 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 20 cm.

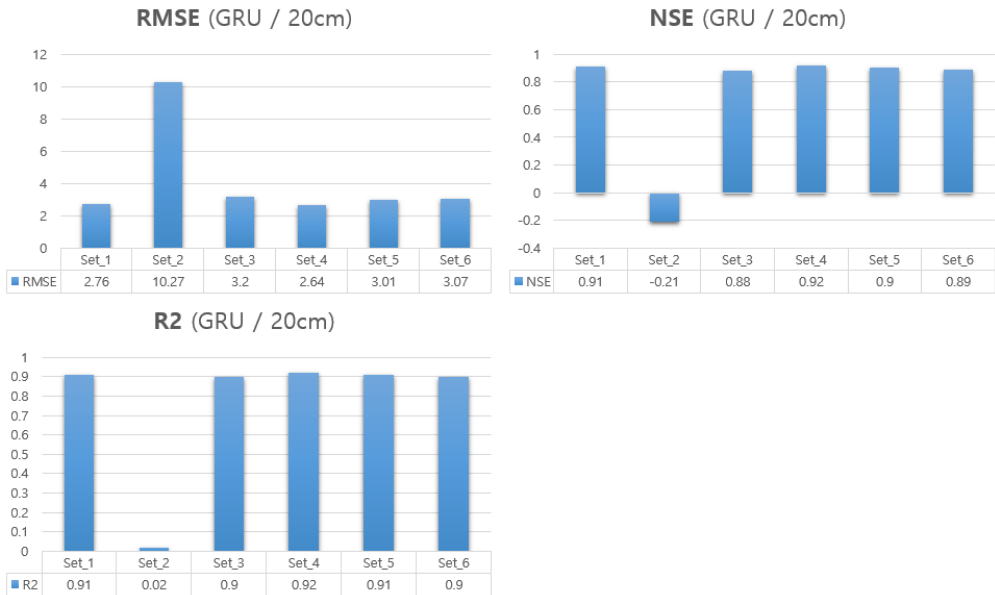


Figure 26 RMSE, NSE, and R2 values of GRU model predictions at a depth of 20 cm.

20cm 깊이에서 각 모델 별 예측값을 평가 지표를 통해 평가하였을 시 10cm 깊이 예서와 마찬가지로 Set\_2의 정확도가 큰 차이로 가장 낮다는 것을 확인할 수 있었고, 다만 10cm 깊이에서와는 다르게 Set\_4의 정확도가 모든 모델과 지표에서 Set\_1의 정확도 보다 높은 것을 확인할 수 있었다. 또한 Set\_3, Set\_5, Set\_6 정확도의 순서를 정의하기 어려웠던 10cm 깊이에서와는 다르게 Set\_5, Set\_6, Set\_3 순서대로 정확도가 높은 것이 확인되었다.

실험 결과를 통해 판단할 수 있는 사실은 20cm 깊이에서 인공지능을 활용하여 지온 데이터를 예측할 때 기온이 가장 큰 영향력을 미치고 습도, 일사량, 풍속의 순으로 영향력이 미친다는 것이다. 향의 경우 예측 정확도를 낮춰 예측에 방해가 된다는 것 또한 판단할 수 있다.

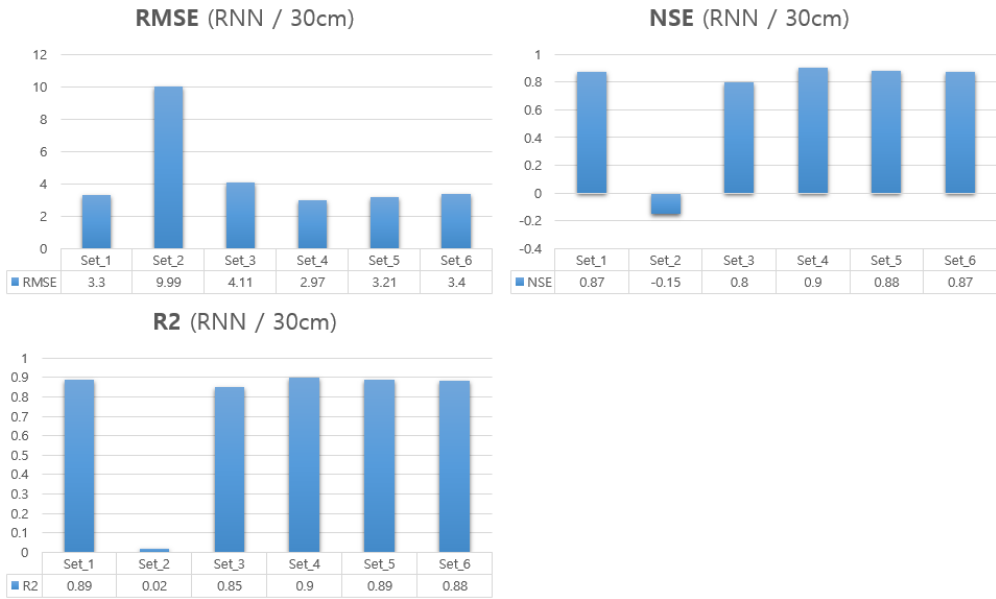


Figure 27 RMSE, NSE, and R2 values of RNN model predictions at a depth of 30 cm.

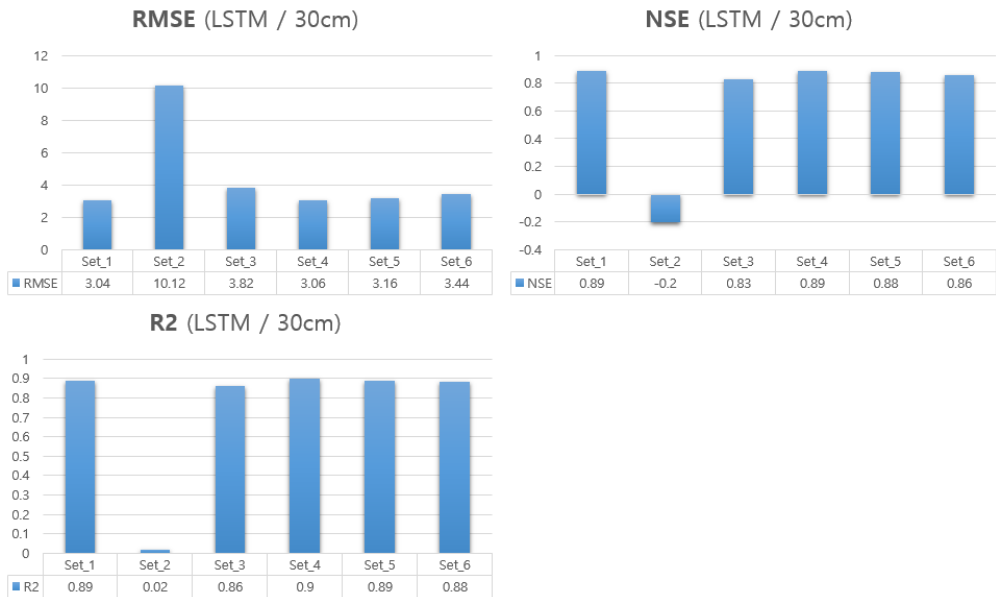


Figure 28 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 30 cm

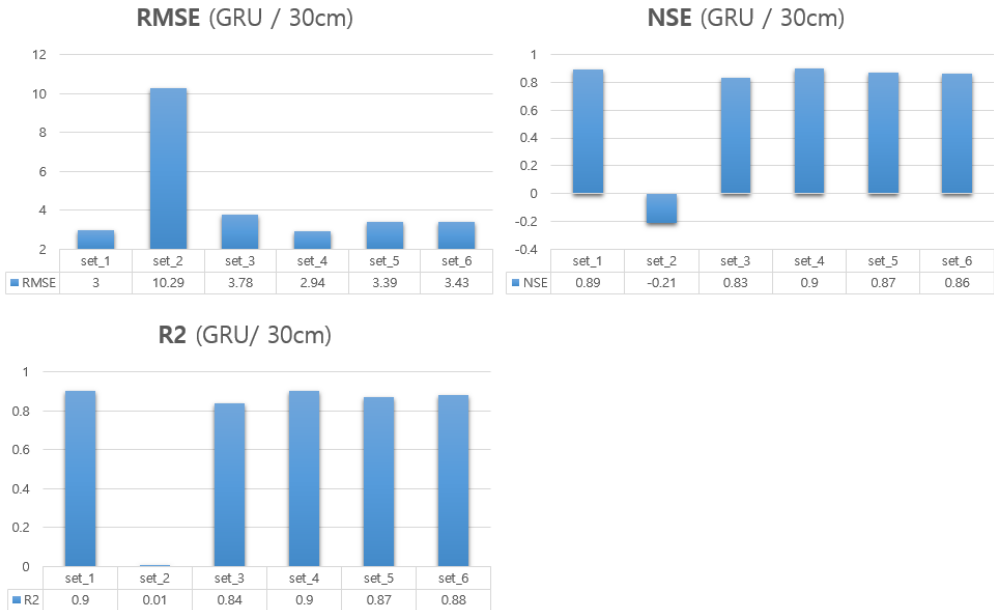


Figure 29 RMSE, NSE, and R2 values of GRU model predictions at a depth of 30 cm

30cm 깊이에서도 각 모델별 예측값을 평가 지표를 통해 평가하였을 시 Set\_4, Set\_1, Set\_5, Set\_6, Set\_3, Set\_2의 순서대로 정확도가 높은 것을 확인하였다. 실험 결과를 통해 판단할 수 있는 사실은 다음과 같다.

30cm 깊이에서 인공지능을 활용하여 지온 데이터를 예측할 때 기온이 가장 큰 영향력을 미치며, 습도, 일사량, 풍속의 순으로 영향력이 미친다. 풍향의 경우 예측 정확도를 낮춰 예측에 방해가 된다.

각 모델별 깊이별 정확도를 평가하였을 때 산점도 그래프와 마찬가지로 기상인자 중 기온을 제외하고 모델을 만들었던 Set\_2의 데이터 세트를 이용한 예측이 모든 깊이와 모델에서 정확도가 대폭 감소하였다. 이러한 결과로 보아 깊이별 지온 데이터 예측에 기온 데이터는 다른 인자들에 비해 영향력이 크다고 판단할 수 있다. 또한 풍향을 제외하고 모델을 만들었던 Set\_4의 데이터 세트 구성을 활용한 예측 모델은 대부분의 깊이와 모델에서 모든 기상인자를 사용하였던 Set\_1의 데이터 세트를 활용한 예측 모델



보다 높은 정확도를 나타내는 것으로 확인되었으며, 이를 통해 풍향은 기상인자와 인공지능을 활용하여 지온을 예측하는 모델을 만들 시 예측에 방해되는 인자라고 판단할 수 있다.

따라서 이후 실험에서는 풍향은 지온 예측에 필요한 기상 인자로써 제외하도록 하였다.

#### 4.1.2. 전진 선택법을 통한 기상인자 최적의 매개변수 조합 선택

풍향을 제외한 네 가지의 기상인자 중 가장 성능이 좋은 최적의 매개변수 조합을 선택하기 위해 전진 선택법(Feedforward Selection)을 사용하였다. 전진 선택법이란 단계적 회귀분석 방법 중 하나로써 설명력이 가장 큰 매개변수부터 하나씩 추가하여 정확도의 증감을 확인하는 방법이다. 한번 선택된 변수는 제거되지 않아 최적의 입력 변수 집합이 선택되지 않을 수 있다는 단점이 있지만 이러한 단점을 보완하기 위해 앞서 매개변수를 하나씩 소거하며 영향력의 여부와 영향력 크기의 순서를 확인한 바 있다. 적은 계산량을 통해 결과를 얻을 수 있다는 장점이 있다.

Table 6 Parameters of each data set

data set	Parameters
Set_7	air temperature
Set_8	air temperature / humidity
Set_9	air temperature / humidity / insolation

최적의 매개변수 조합을 선택하기 위해 Table 6 과같이 매개변수 조합을 구성하였다. Set\_9에 풍속이 추가된 데이터 세트는 Set\_4와 동일하기 때문에 Set\_4의 결과값을 추가하여 비교하였다. 세 종류의 머신러닝 모델과 세 개의 깊이에서 예측을 진행하였으며, 각 깊이별, 모델별 실측값-예측값 산점도 그래프는 다음과 같다.

10cm

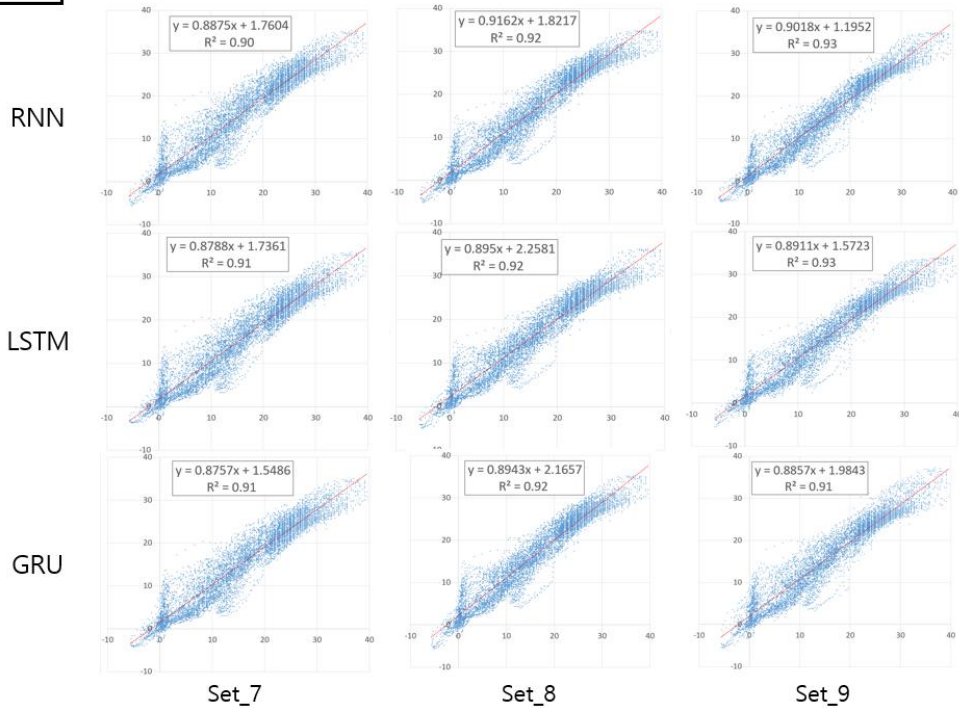


Figure 30 Scatterplot graph of each set of RNN, LSTM, GRU models at a depth of 10 cm

10cm 깊이에서 각 데이터 세트들의 산점도 그래프는 모든 모델에서 선형을 보이며  $R^2$  값이 0.85 이상으로 높은 정확도를 보이는 것을 확인할 수 있다.  $R^2$  값의 경우 단일 입력 변수를 활용한 set\_7에서 변수가 추가될수록 증가하는 것 또한 확인할 수 있었지만 그 차이는 크지 않았기 때문에 평가 지표 수치의 비교를 통해 매개변수 추가에 따른 성능이 평가되어야 할 필요가 있다는 것을 확인하였다.

20cm

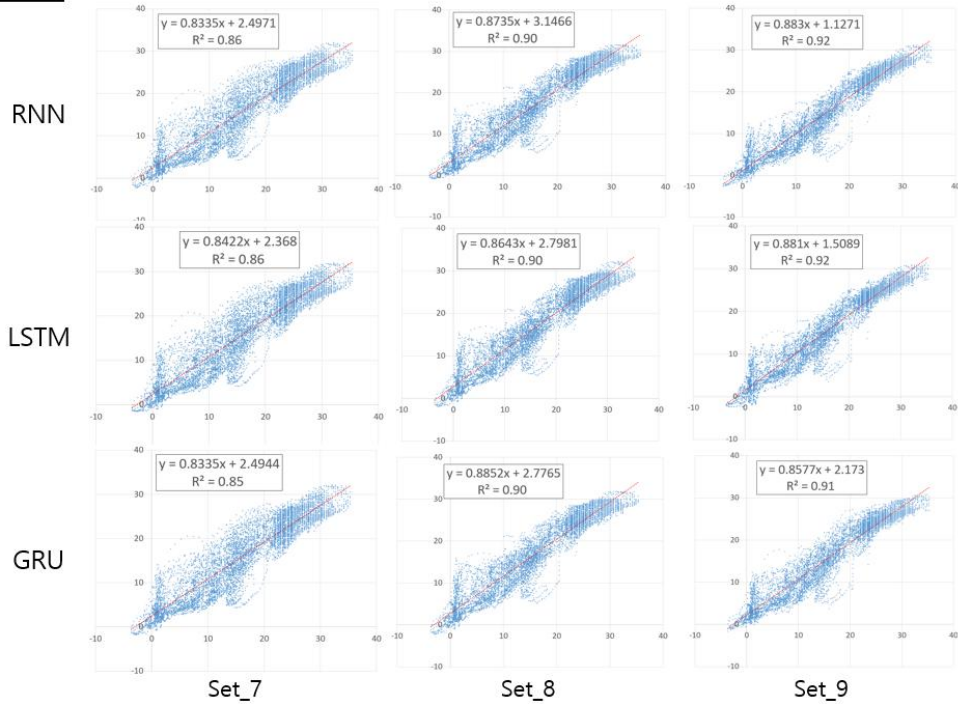


Figure 31 Scatterplot graph of each set of RNN, LSTM, GRU models at a depth of 20 cm

20cm 깊이에서 각 데이터 세트들의 산점도 그래프는 모든 모델에서 선형을 보였고  $R^2$ 값의 경우에도 모든 모델과 데이터 세트에서 0.85 이상으로 높은 정확도를 보이는 것을 확인할 수 있었지만 단일 매개변수를 활용했던 set\_7의 경우 10cm 깊이에서의 예측보다 비교적 크게 감소하였다는 것을 확인할 수 있다. 그뿐만 아니라 매개변수로 습도를 추가했던 set\_8의 예측값과 차이가 크게 나타나는 것을 확인할 수 있다.

30cm

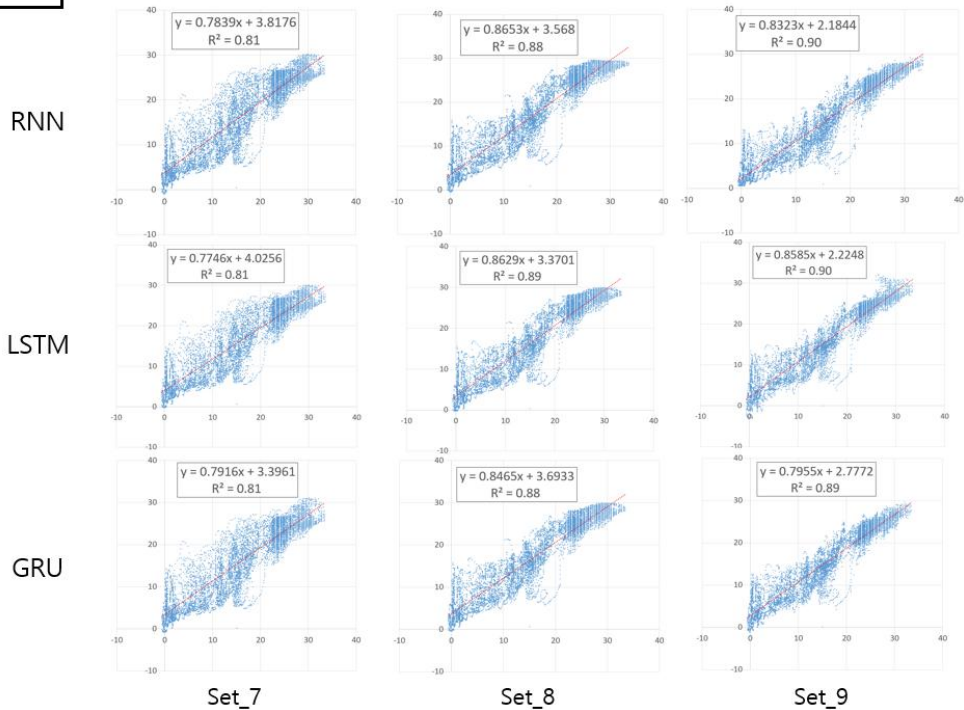


Figure 32 Scatterplot graph of each set of RNN, LSTM, GRU models at a depth of 30 cm

30cm 깊이에서의 산점도 그래프 역시 앞선 그래프들과 마찬가지로 모든 모델과 세트에서 선형을 보였다. 하지만  $R^2$  값의 경우 set\_7에서 0.85 이하의 비교적 낮은 정확도를 보였고, 이는 20cm에서의  $R^2$ 에 비해 많이 감소한 수치이다. 다른 세트에서도 역시 깊이에 따라  $R^2$ 이 감소하는 것을 확인할 수 있었다. 각 깊이별, 모델별 평가 지표를 비교한 그래프는 다음과 같다.

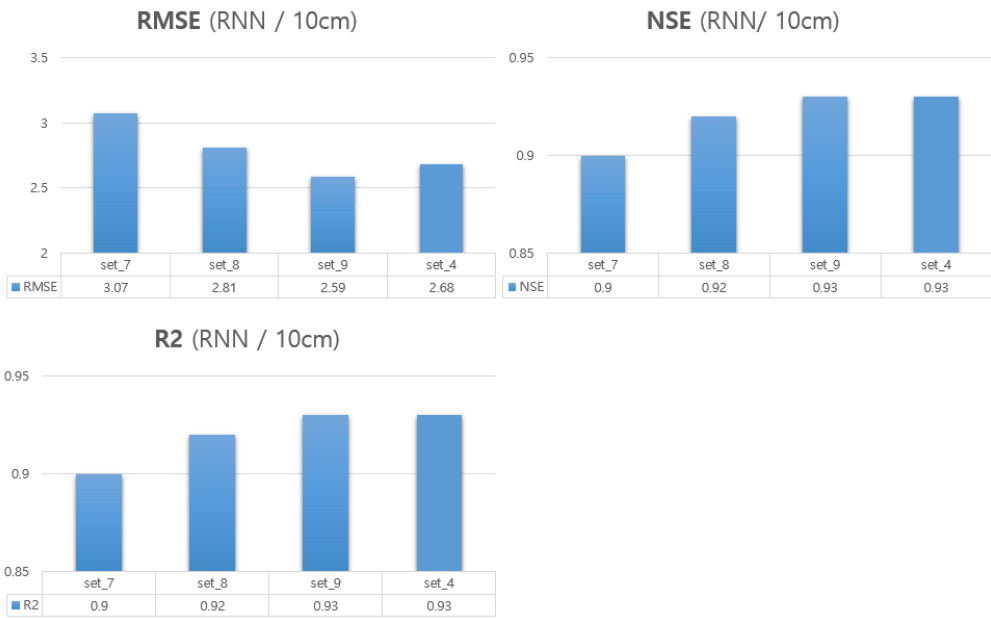


Figure 33 RMSE, NSE, and R2 values of RNN model predictions at a depth of 10 cm

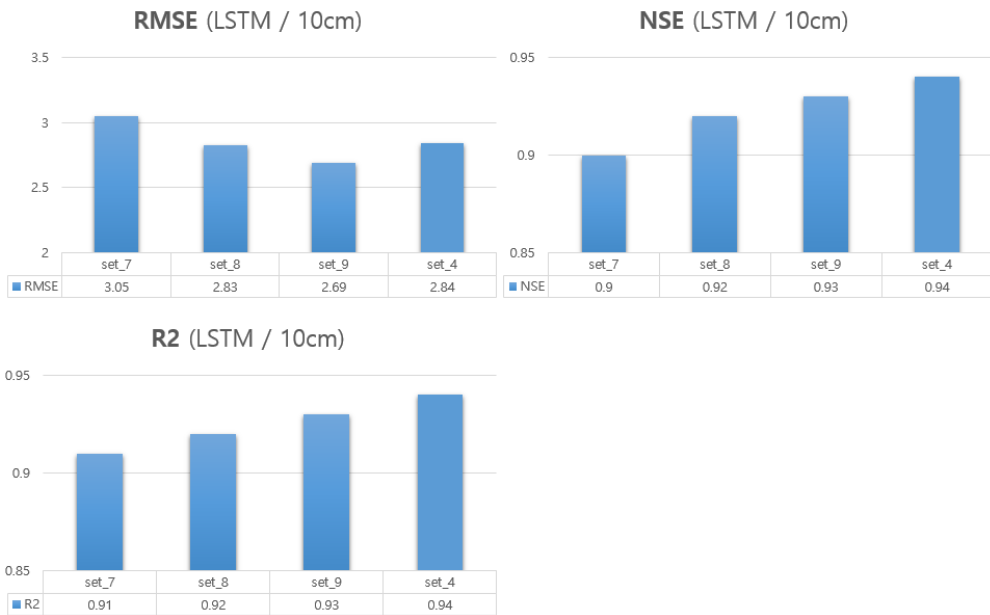


Figure 34 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 10 cm

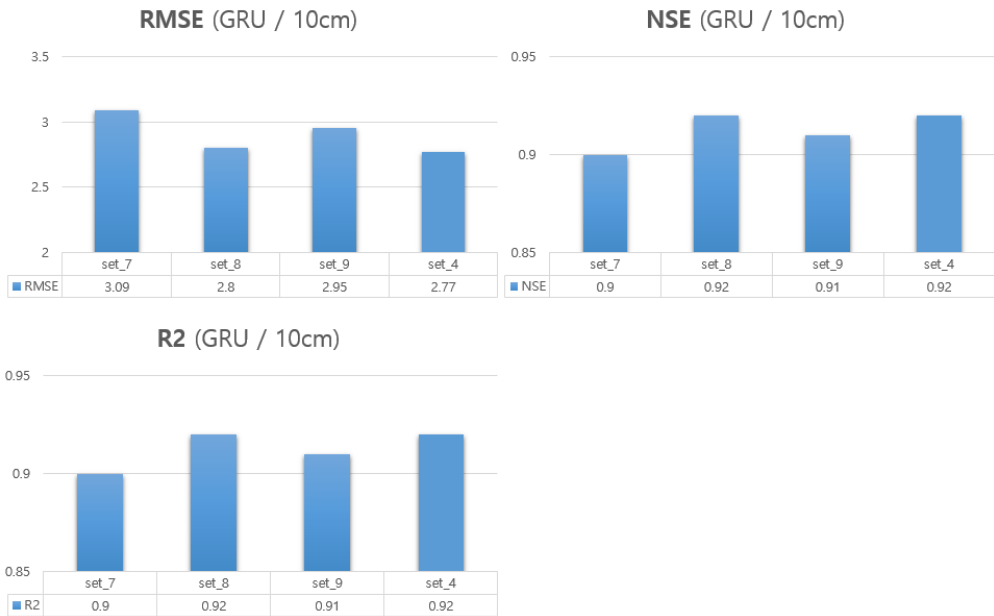


Figure 35 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 10 cm

10cm 깊이에서 각 지표의 수치를 통해 변수를 추가할수록 정확도가 상승하는 추세를 보였지만 그 반대의 경향을 보인 몇몇 구간이 있었다. RNN 모델을 활용한 예측의 경우 Set\_9의 RMSE 값이 2.59°C로 풍속을 매개변수로써 추가한 Set\_4의 RMSE 값인 2.68°C보다 0.9°C가량 낮은 결과가 나타났다. LSTM 모델을 활용한 예측에서도 Set\_9의 RMSE 2.69°C로 Set\_4의 RMSE 값인 2.84°C보다 1.5°C가량 낮은 결과가 나왔다. 다만 이 경우 NSE 값과 R<sup>2</sup> 값은 그 반대의 경향이 나타났다. GRU 모델을 활용한 예측의 경우 Set\_8에서의 모든 수치가 일사량을 매개변수로써 추가한 Set\_9보다 정확도가 높게 나왔다. 하지만 몇몇의 경우를 제외하고는 매개변수가 추가될수록 정확도가 높게 나오는 것으로 확인되었다.

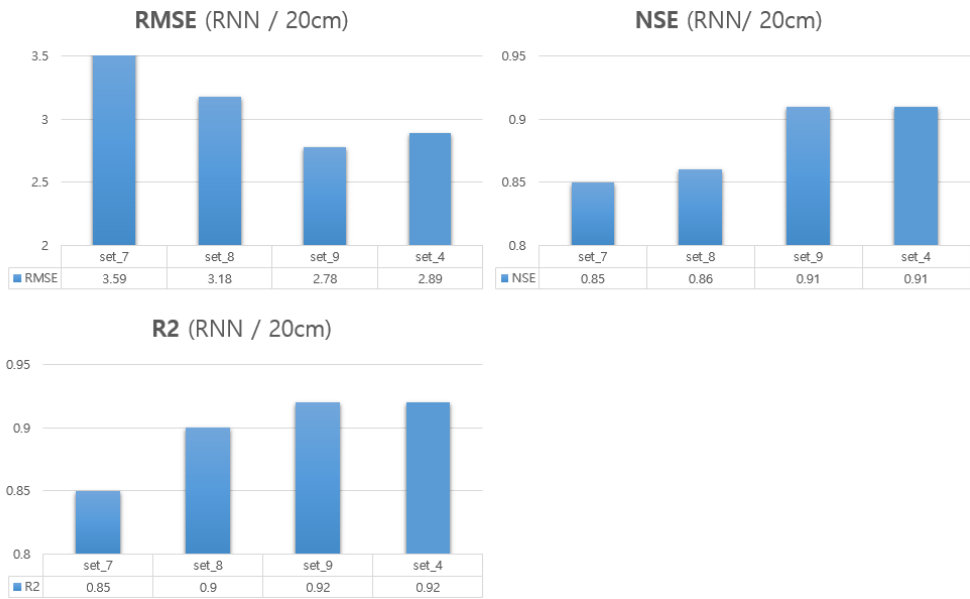


Figure 36 RMSE, NSE, and R2 values of RNN model predictions at a depth of 20 cm

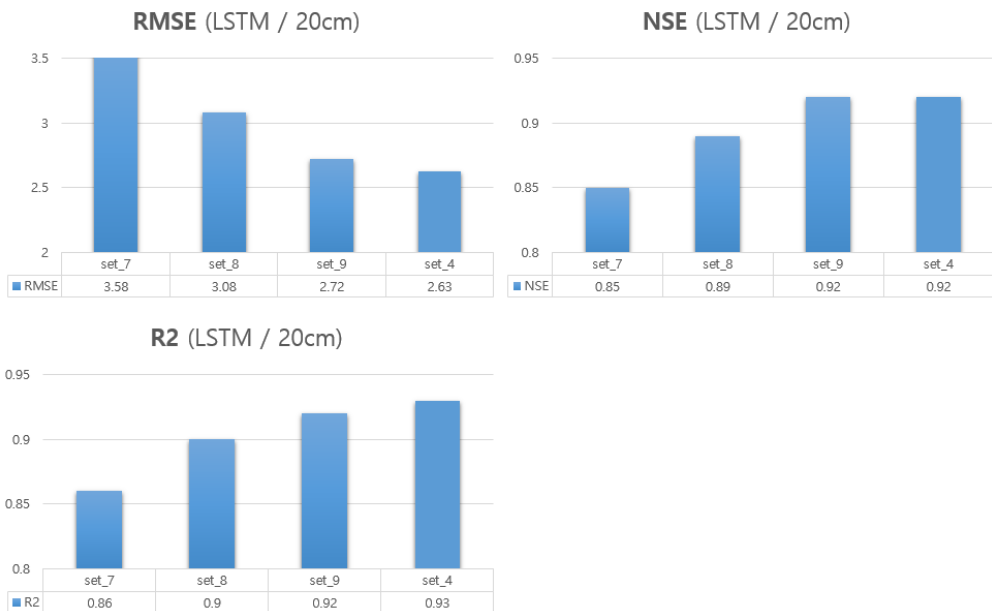


Figure 37 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 20 cm

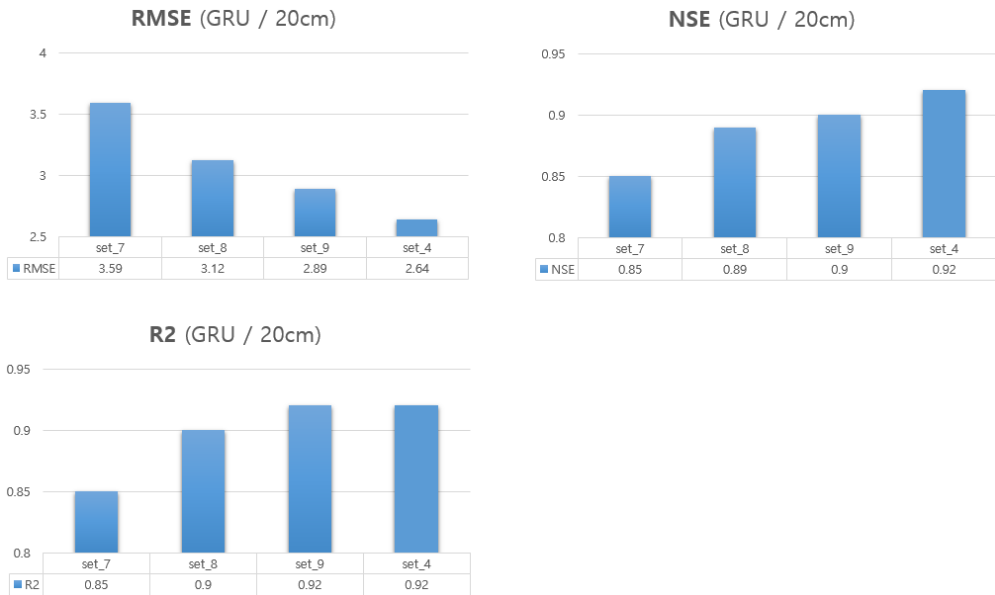


Figure 38 RMSE, NSE, and R2 values of GRU model predictions at a depth of 20 cm

20cm 깊이에서 각 지표의 수치를 통해 변수를 추가할수록 정확도가 상승하는 추세가 10cm 깊이에서의 예측보다 더 뚜렷하게 나타남을 확인할 수 있다. RNN 모델을 활용한 예측에서 set\_9의 RMSE 값이 2.78°C로 set\_4의 RMSE 값인 2.89°C보다 1.1°C가량 낮은 결과가 나온 경우를 제외하고는 매개변수가 추가될수록 높거나 같은 수치의 정확도가 나타는 것을 확인할 수 있다. 이는 산점도 그래프와 마찬가지로 깊이에 따라서 매개변수가 추가될수록 정확도가 더 상승하는 것을 확인할 수 있다. 그뿐만 아니라 지온을 단일 매개변수로 활용한 Set\_7의 경우 모든 수치에서 10cm 깊이에서의 예측보다 정확도가 비교적 큰 폭으로 감소하는 것을 확인할 수 있다. 이는 깊이에 따라 기온의 영향력이 크게 감소한다는 Alizamir(2019)의 연구 결과를 뒷받침할 수 있는 결과이다.



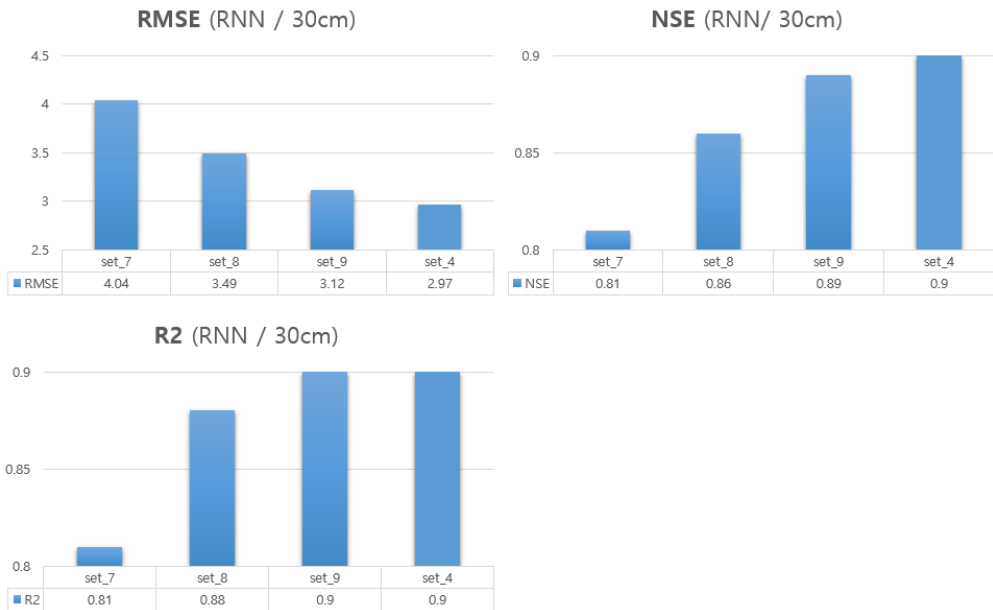


Figure 39 RMSE, NSE, and R2 values of RNN model predictions at a depth of 30 cm



Figure 40 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 30 cm

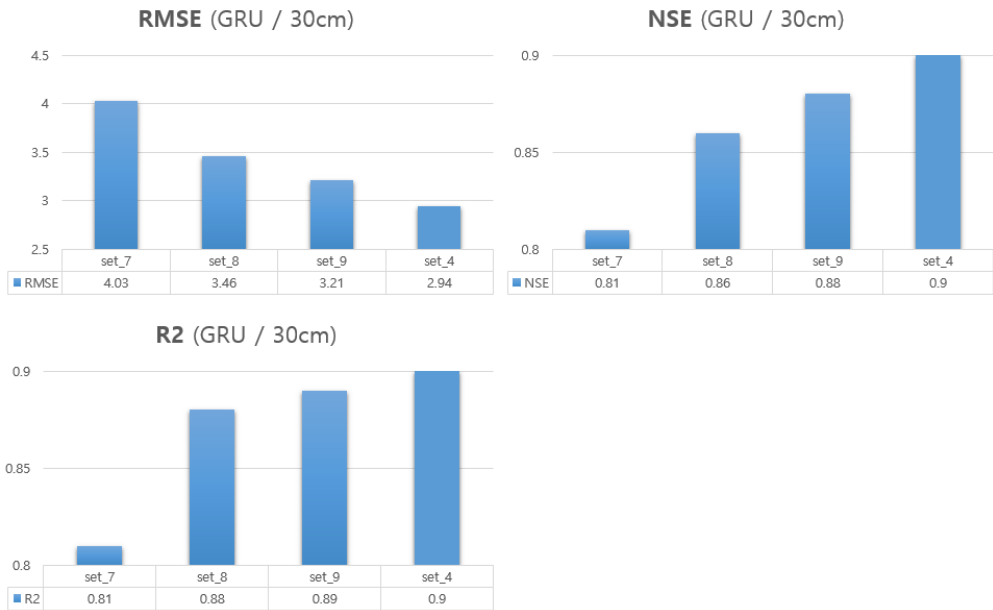


Figure 41 RMSE, NSE, and R2 values of GRU model predictions at a depth of 30 cm

30cm 깊이에서 각 지표의 수치를 통해 변수를 추가할수록 정확도가 상승하는 추세가 역시 20cm 깊이에서의 예측보다 더 뚜렷하게 나타남을 확인할 수 있었다. LSTM 모델을 활용한 예측에서 set\_9의 RMSE 값이 3.02°C로 set\_4의 RMSE 값인 3.06°C보다 0.04°C가량 낮은 결과가 나왔는데, 이는 사실상 같은 수치라고 봐도 무방하다. 이 경우를 제외하고는 모든 구간에서 매개변수가 추가될수록 높거나 같은 수치의 정확도가 나타나는 것이 확인되었다. 지온을 단일 매개변수로 활용한 Set\_7의 경우 모든 수치에서 감소 폭이 증가하였다.

진진선택법을 통한 각 기상인자의 영향력 평가에서 기온, 습도, 풍속, 일사량 데이터 모두 활용한 Set\_4 데이터세트의 예측모델이 다른 예측모델에 비해 성능이 좋은 것으로 평가된다. 단일 매개변수를 활용한 Set\_7은 가장 정확도가 낮았으며, 깊이에 따라 감소되는 정확도가 다른 데이터 세트보다 큰 양상을 보였다. 이는 지온에 미치는 기온의 영향력이 깊어질

수록 감소했던 앞전 실험들과 같은 양상이다. 가장 영향력이 큰 기온을 제외한 나머지 기상인자 역시 깊이별 지온을 예측하는데 있어 영향을 끼치며, 깊이가 증가함에 따라 예측에 있어 필요성을 확인하였다. 특히 습도의 경우 물은 토양입자에 비해 비열이 낮아 지표면에서 교환된 열 에너지의 이송이 빠르고 깊은 깊이까지 도달이 가능하기 때문에 깊이가 깊어지더라도 기온에 비해 영향력의 감소가 적을 것으로 판단된다.

## 4.2. 주기성 변수

Table 7 Parameters of each data set

set	Parameters
Set_10	air temperature / humidity / wind speed / insolation / The average monthly temperature nationwide
Set_11	air temperature / humidity / wind speed / insolation / altitude of the sun
Set_12	air temperature / humidity / wind speed / insolation / Soil temperature by depth in other region

주기성 변수가 예측 모델의 정확도에 미치는 영향을 확인하기 위해 Table 7 과같이 매개변수 조합을 구성하고 세 종류의 머신러닝 모델과 세 개의 깊이에서 예측을 진행하였다. 각 깊이별, 모델별 실측값-예측값 산점도 그래프는 다음과 같다.

10cm

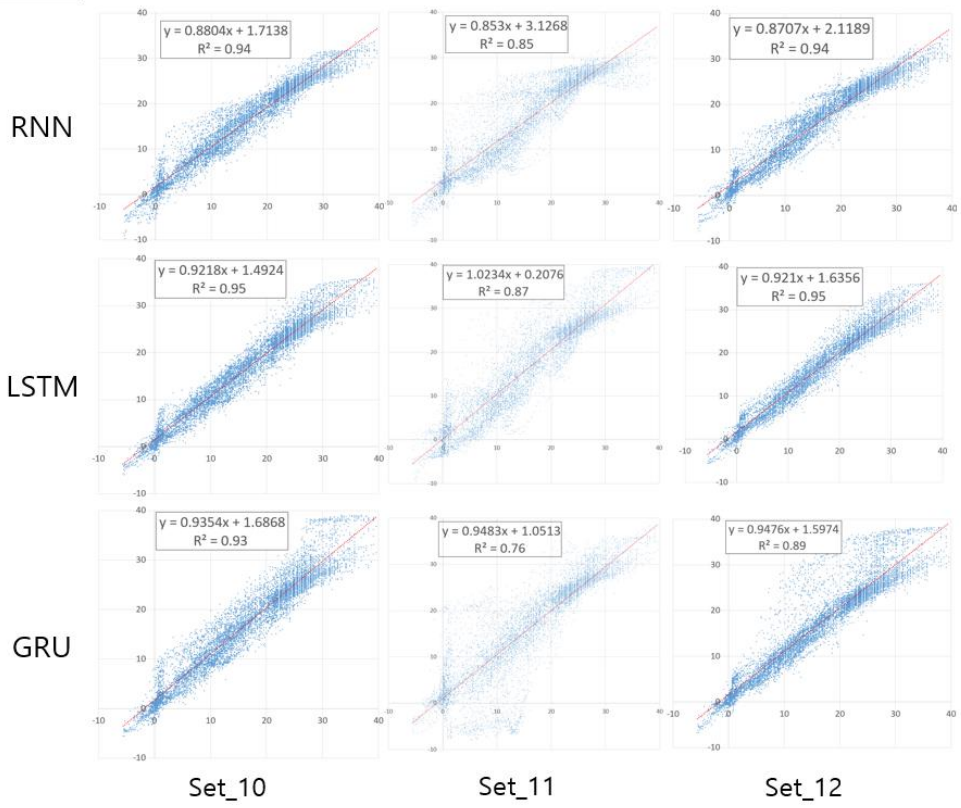


Figure 42 Scatter plot graph of each set of RNN, LSTM, GRU models at a depth of 10 cm

10cm 깊이에서 산점도 그래프를 확인한 결과 set\_11의 경우 직선형태의 그래프를 나타내고 있으나, 산점도가 비교적 분산되어 있는 형태이며,  $R^2$ 이 비교적 낮은 것으로 확인된다. 반면에 set\_10과 set\_12의 경우 산점도가 분산되어있지 않고 대부분의  $R^2$ 값이 이 set\_4보다 높게 나오는 것을 확인할 수 있다.

20cm

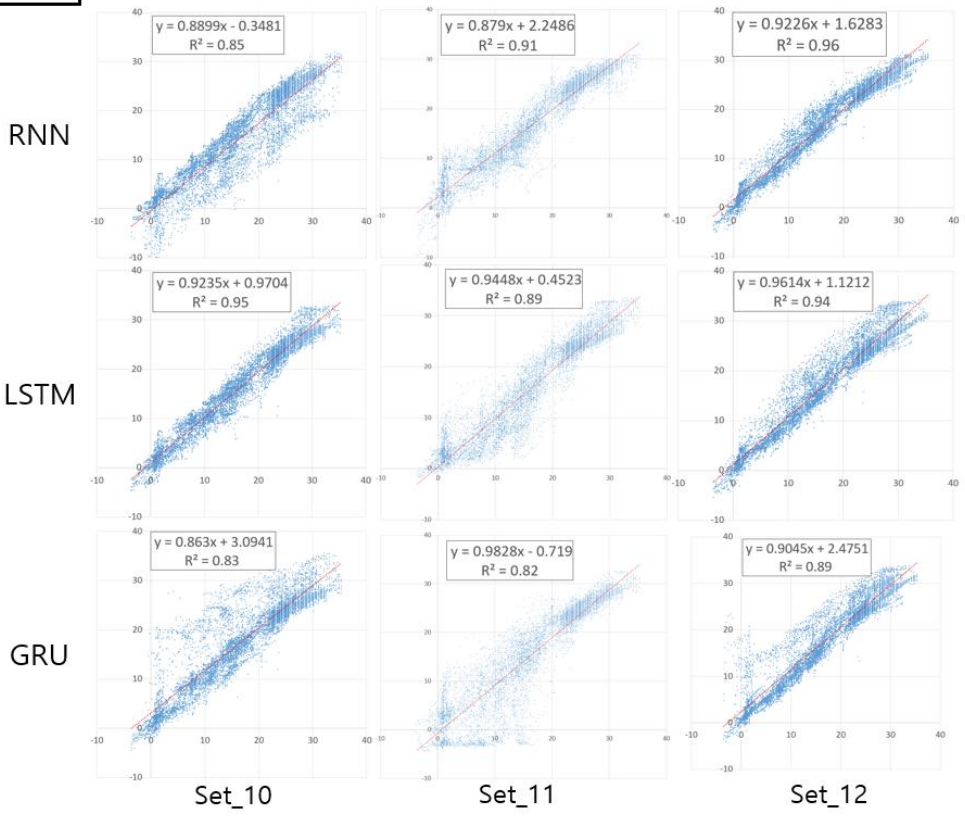


Figure 43 Scatter plot graph of each set of RNN, LSTM, GRU models at a depth of 20 cm

20cm 깊이에서는 Set\_10과 Set\_11이 모델에 따라  $R^2$ 값의 차이가 큰 것으로 확인된다. Set\_10의 경우 LSTM를 활용한 예측값의 경우  $R^2$ 은 0.93으로 높은 정확도를 보였지만, 그 외의 모델에서는 0.85이하의 비교적 낮은 정확도를 보였다. Set\_11의 경우 GRU모델에서  $R^2$ 이 0.79로 낮은 수치가 확인되었다. Set\_12는 다른 데이터세트에 비해 모든 모델에서 안정적인 편이었지만, 기상인자만을 매개변수로 활용했던 전 실험들의 예측에 비해서는 모델에 따라 차이가 큰 편이다.

30cm

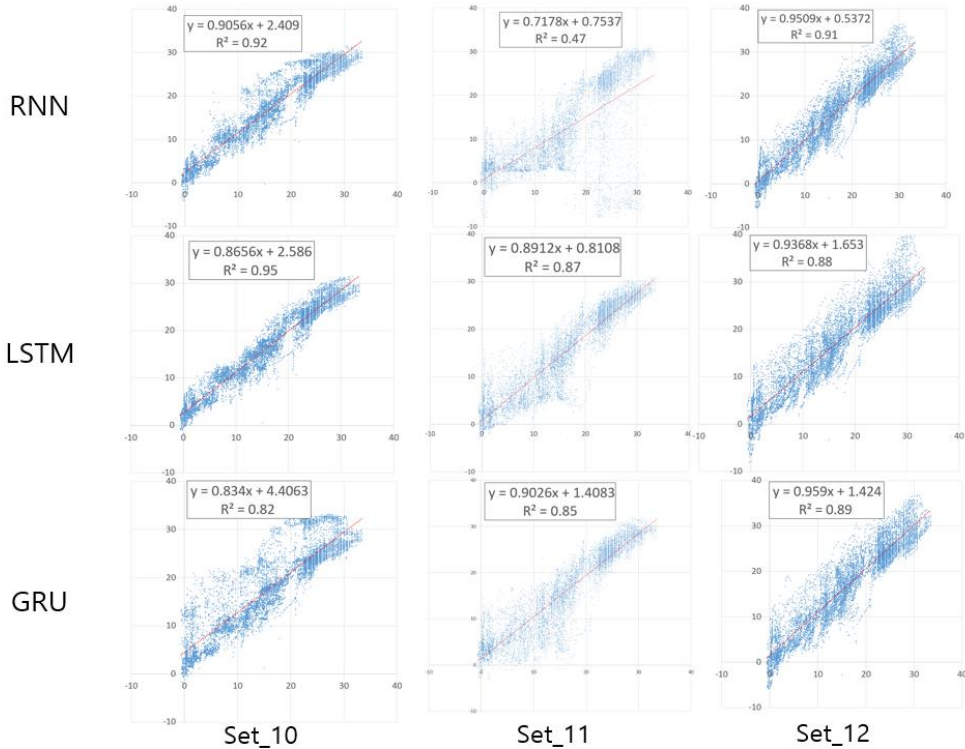


Figure 44 Scatter plot graph of each set of RNN, LSTM, GRU models at a depth of 30 cm

30cm 깊이에서는 모델 따른  $R^2$ 값의 차이가 더 증가하였다. Set\_11의 경우 RNN모델을 활용한 예측의  $R^2$ 값은 0.47로 비교적 매우 낮은수치가 확인되었다. Set\_10의 경우 LSTM을 활용한 예측값 에서의  $R^2$ 값은 0.95로 매우 높은 수치가 확인되었지만, GRU를 활용한 예측값 에서의  $R^2$ 값은 0.82로 비교적 낮은 수치가 확인 되었다. 각 깊이별, 모델별 평가지표를 비교한 그래프는 다음과 같다.

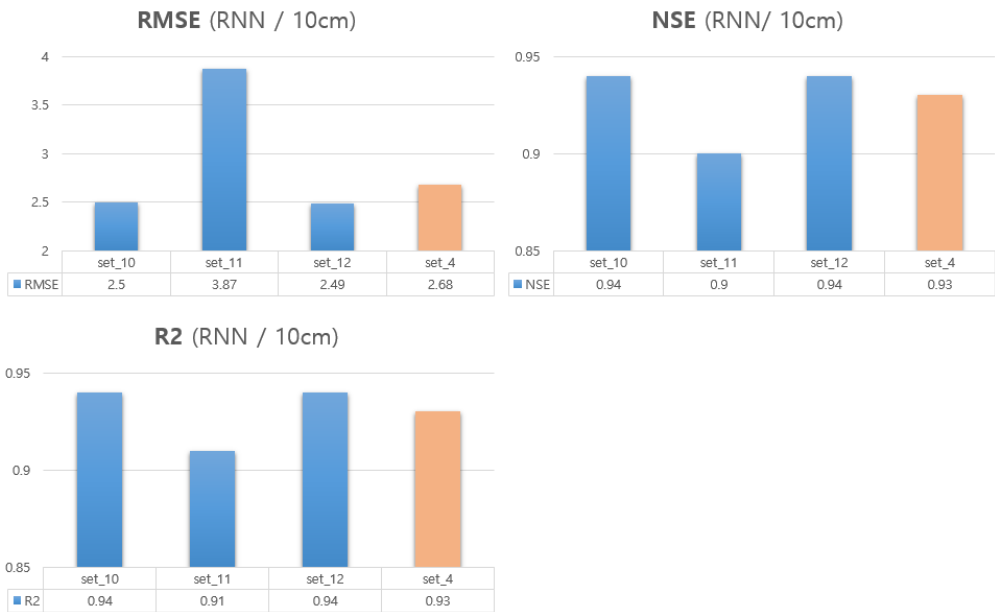


Figure 45 RMSE, NSE, and R2 values of RNN model predictions at a depth of 10 cm

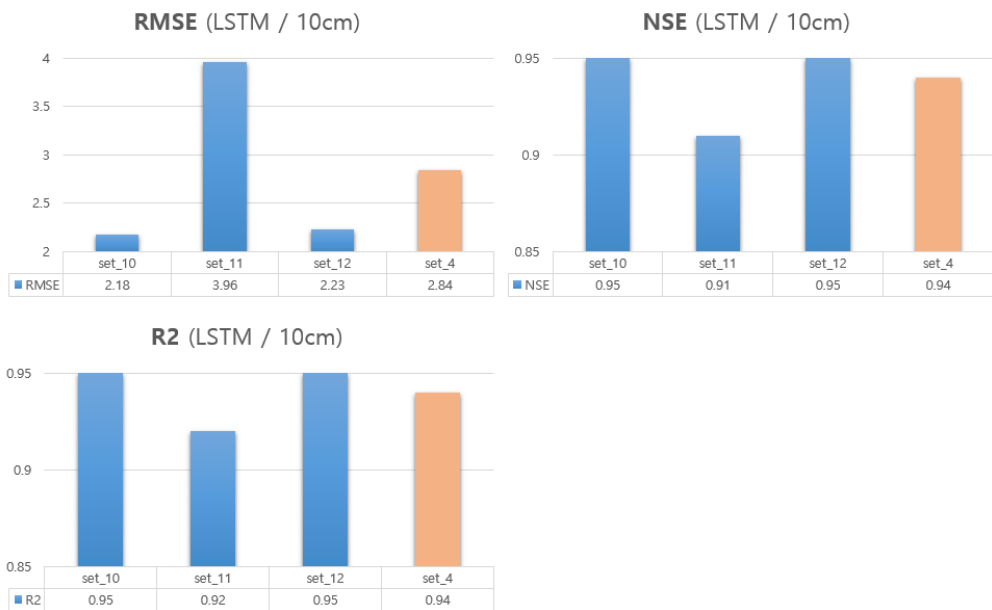


Figure 46 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 10 cm





Figure 47 RMSE, NSE, and R2 values of GRU model predictions at a depth of 10 cm

10cm 깊이에서 주기성 변수를 매개변수로 활용한 예측값 들 은 각자 다른 양상을 보였다. Set\_10은 모든 모델에서 Set\_4 보다 높은 정확도를 보이며, Set\_11은 경우 모든 모델에서 Set\_4 보다 낮은 정확도를 보였다. Set\_12의 경우 RNN 모델과 LSTM 모델을 활용한 예측값은 Set\_4보다 높은 정확도를 보였지만 GRU 모델을 활용한 예측값은 Set\_4보다 낮은 정확도를 보이는 것으로 확인되었다.

결과에 따르면 10cm 깊이에서 전국 월평균 기온과, 타지역 지온 값은 정확도를 상승시키는데 영향을 끼치는 인자이며, 시간에 따른 태양고도는 예측에 방해가 되는 인자라고 사료될 수 있다.

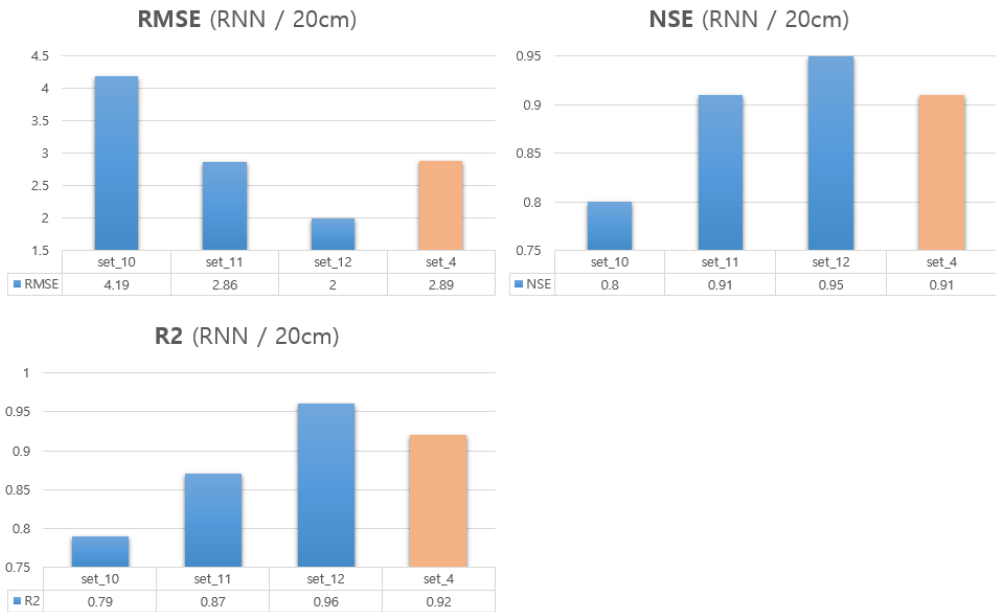


Figure 48 RMSE, NSE, and R2 values of RNN model predictions at a depth of 20 cm



Figure 49 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 20 cm

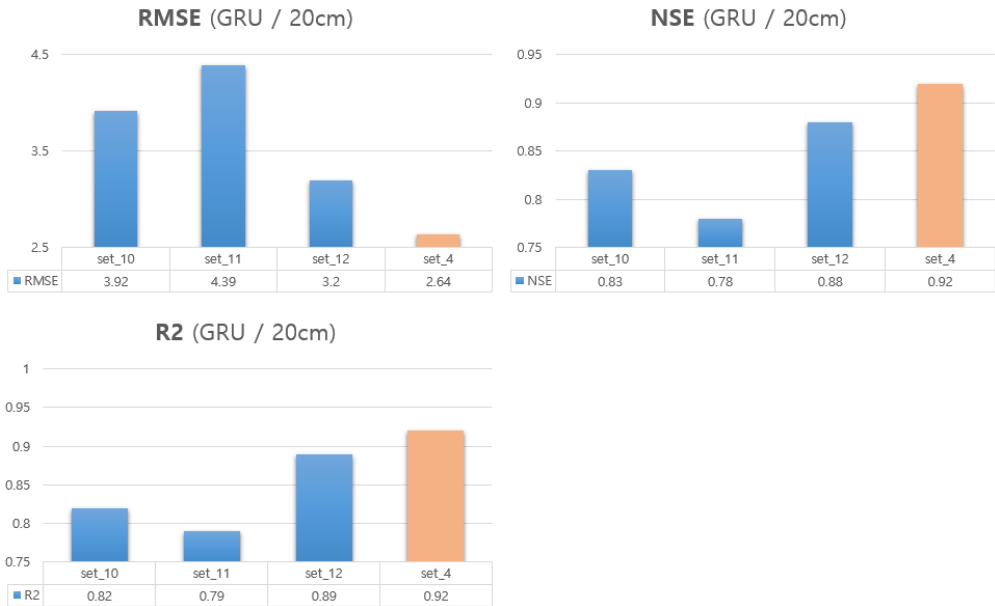


Figure 50 RMSE, NSE, and R2 values of GRU model predictions at a depth of 20 cm

20cm 깊이에서는 10cm 깊이에서의 예측과 다른 양상을 보였는데, Set\_10의 경우 LSTM 모델을 제외한 나머지 모델에서의 예측값 모두 Set\_4의 예측값보다 정확도가 떨어졌다. 다만 LSTM 모델을 활용했을 시에는 Set\_4의 예측보다 정확도가 높았다. Set\_11의 경우에는 RNN 모델을 활용했을 시에는 Set\_4의 예측과 비슷한 정확도를 보였으며, 나머지 두 모델을 활용했을 시에는 Set\_4의 예측에 비해 정확도가 낮았다. 20cm 깊이에서도 시간에 따른 태양고도의 변화량을 활용한 매개변수가 예측에 방해가 되는 10cm 깊이에서의 예측과 같은 경향을 보였다. Set\_12의 경우 GRU 모델을 활용했을 시 Set\_4의 예측보다 정확도가 낮았지만 다른 두 모델을 활용했을 시에는 Set\_4의 예측보다 높은 정확도를 보였다.

20cm 깊이에서는 모델에 따라 매개변수가 정확도에 미치는 영향력의 차이가 10cm 깊이보다 증가하였다. 또한 마찬가지로 태양 고도를 매개변수로 활용한 Set\_11의 데이터세트는 전체적으로 Set\_4의 데이터 세트보다 성능이 떨어지는 데이터 세트로 확인되었다.

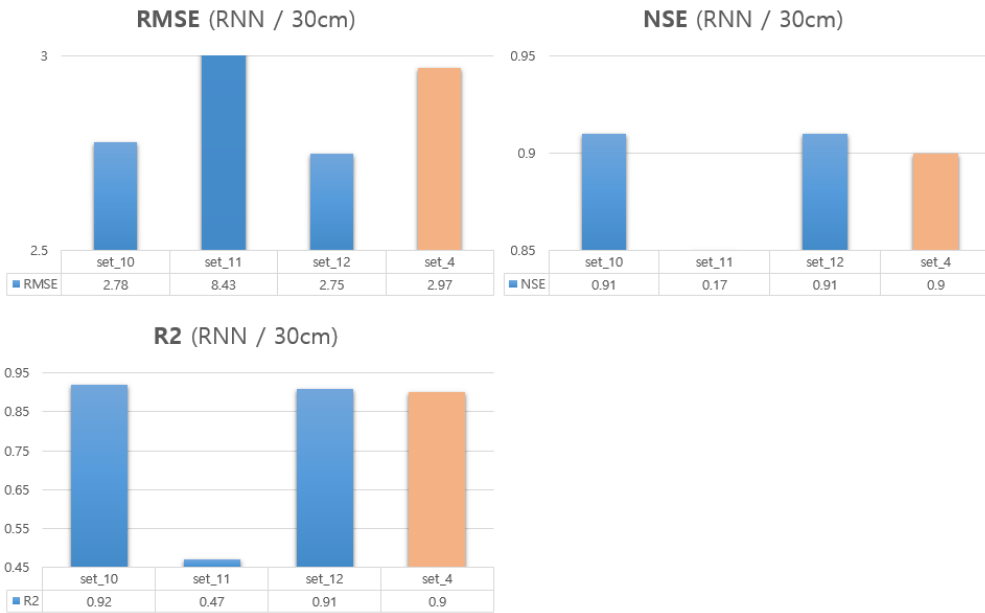


Figure 51 RMSE, NSE, and R2 values of RNN model predictions at a depth of 30 cm



Figure 52 RMSE, NSE, and R2 values of LSTM model predictions at a depth of 30 cm



Figure 53 RMSE, NSE, and R2 values of GRU model predictions at a depth of 30 cm

30cm 깊이에서는 모델에 따른 정확도의 편차가 급증한 것을 확인할 수 있다. Set\_10의 경우 RNN 모델과 LSTM 모델을 활용한 예측에서는 Set\_4보다 정확도가 높았지만 GRU 모델을 활용한 예측에서는 Set\_4보다 정확도가 떨어지는 것을 확인할 수 있다. Set\_11의 경우에는 30cm 깊이에서 역시 정확도가 떨어지는 것으로 보아 시간에 따른 태양고도의 변화량을 이용한 매개변수는 깊이별 지온예측에서 예측을 하는데 방해되는 요소라고 판단된다. Set\_12의 경우에는 RNN 모델을 활용한 예측에서는 Set\_4보다 정확도가 높았지만 다른 두 모델을 활용한 예측에서는 Set\_4보다 정확도가 떨어지는 것을 확인할 수 있다.

주기성을 갖는 인자를 매개변수로 활용했을 시 모델에 따라 예측 정확도에 미치는 영향이 달랐다. 월평균 기온을 주기성 매개변수로 활용했던 Set\_10 데이터세트는 LSTM에서 특히 높은 예측 정확도를 보였고, 타지역 깊이별 지온데이터를 매개변수로 활용했던 Set\_12 데이터 세트의 경우

에는 RNN에서 특히 높은 예측 정확도를 보였다. 반면에 시간에 따른 태  
양고도 값을 매개변수로 활용한 Set\_11은 전체적으로 정확도를 감소시키  
는 요인으로 작용하였다.

Set\_10 데이터세트와 LSTM 모델을 활용한 Set\_10\_LSTM과 Set\_12 데  
이터세트와 RNN 모델을 활용한 Set\_12\_RNN 이 모든 깊이에서 가장 높  
은 정확도를 보였다. 두 예측값 모두 기상인자만 활용했던 모델 중 전체  
적으로 가장 성능이 좋았던 Set\_4\_GRU 보다 높은 예측 정확도를 보였으  
며, 주기성을 갖는 매개변수를 활용했을 시 모델에 따라 성능 향상을 보  
일 수 있음을 확인 하였다.

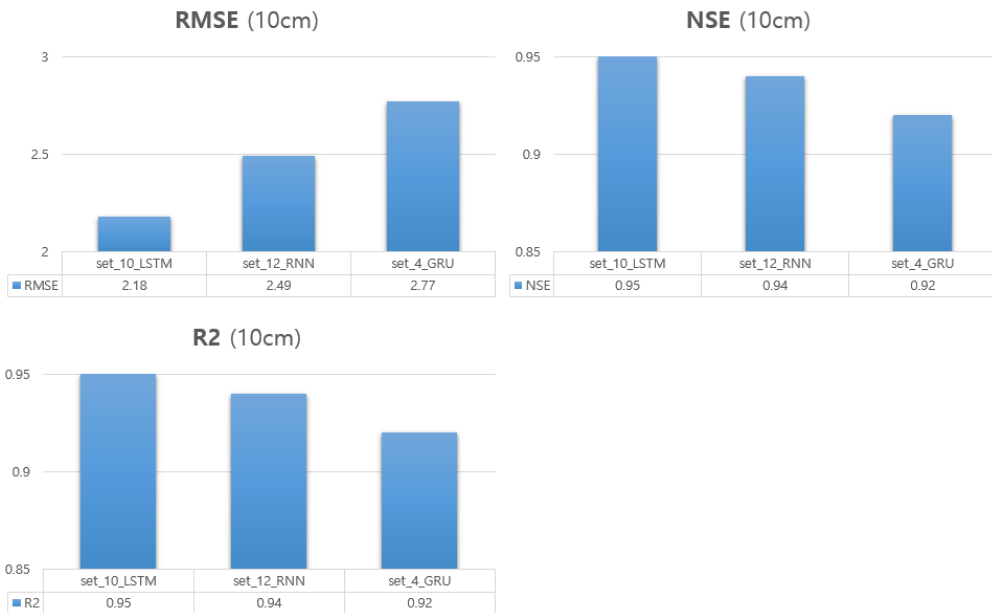


Figure 54 RMSE, NSE, and R2 values for each data set and model combination at a depth of 10 cm.

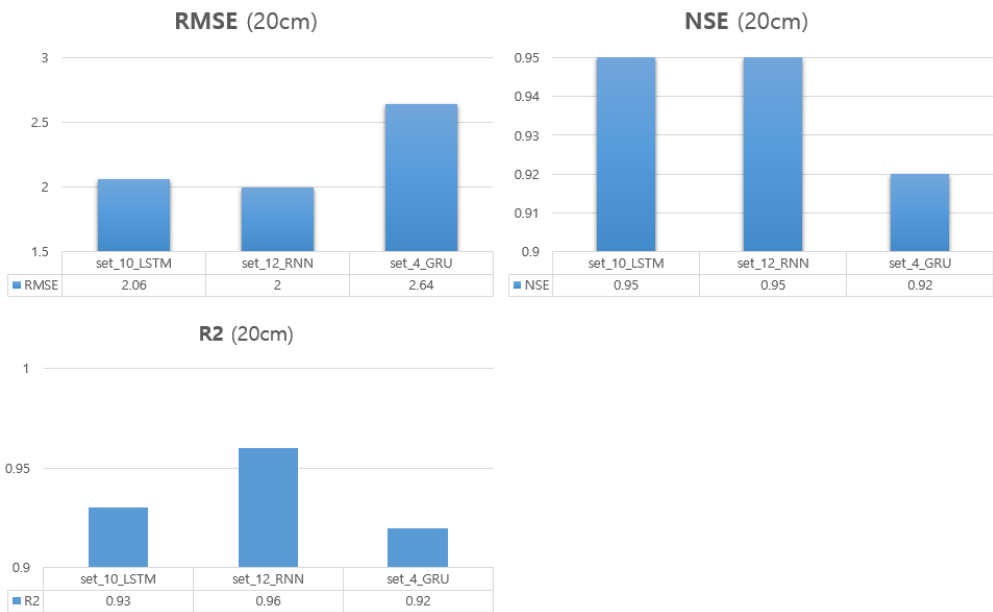


Figure 55 RMSE, NSE, and R2 values for each data set and model combination at a depth of 20 cm.

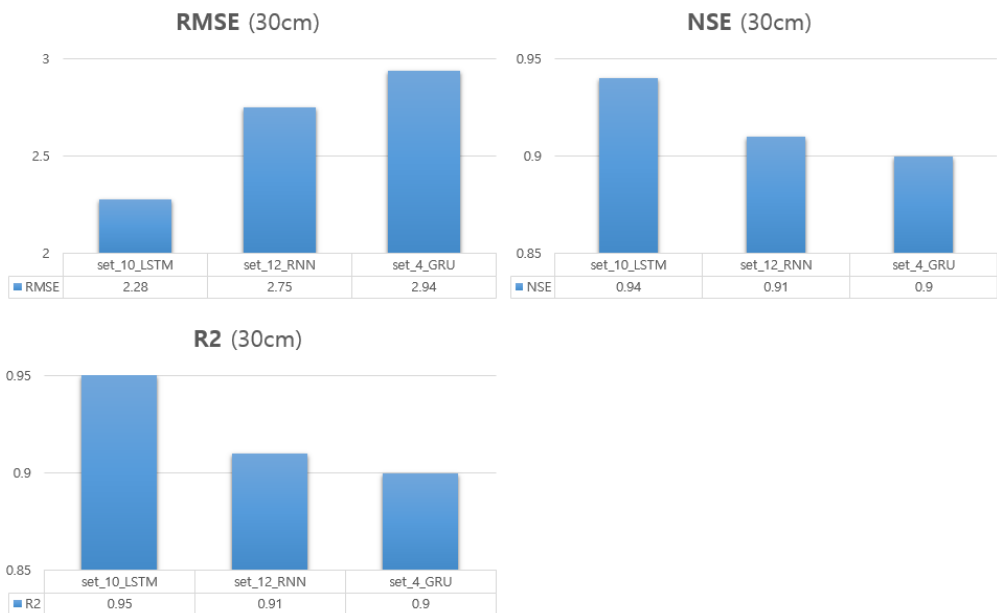


Figure 56 RMSE, NSE, and R2 values for each data set and model combination at a depth of 30 cm.

### 4.3. 데이터 비교 및 분석

실험 중 성능이 가장 좋았던 데이터세트-모델의 예측값 A와 예측값B를 실측값과 비교하고 분석하였다. 여기서 예측값 A는 Set\_10 데이터 세트와 LSTM 모델을 활용하여 예측된 예측값, 예측값 B는 Set\_12 데이터 세트와 RNN 모델을 활용하여 예측된 예측값이다.

먼저 깊이에 따른 연별, 일별 추세를 비교하였고, 각 예측값들과 실측값의 잔차를 확인하고 잔차 원인을 분석하였다.

#### 4.3.1. 추세 비교

실측값과 두 개의 예측값의 유사성을 확인하기 위해 추세를 비교하였다. 깊이에 따라 실험이 진행된 2021년 1월 1일 ~ 2021년 11월 30일 동안의 전체적인 추세와 2021년 8월 15일 하루의 추세를 얼마나 유사성을 가지고 있는지 비교하였다.

11개월의 기간 동안의 연간 전체 추세는 세 구간으로 나뉘었는데 1월 1일 0시~연중 최소 지점, 연중 최소 지점~연중 최고 지점, 연중 최고 지점~11월 30일 23시 세 개의 구간으로 분리하여 추세를 확인하였으며, 8월 15일 하루의 추세는 실측값, 예측값 A, 예측값 B의 데이터를 그래프로 비교하며 유사성을 확인하였다.



### 4.3.1.1. 10cm

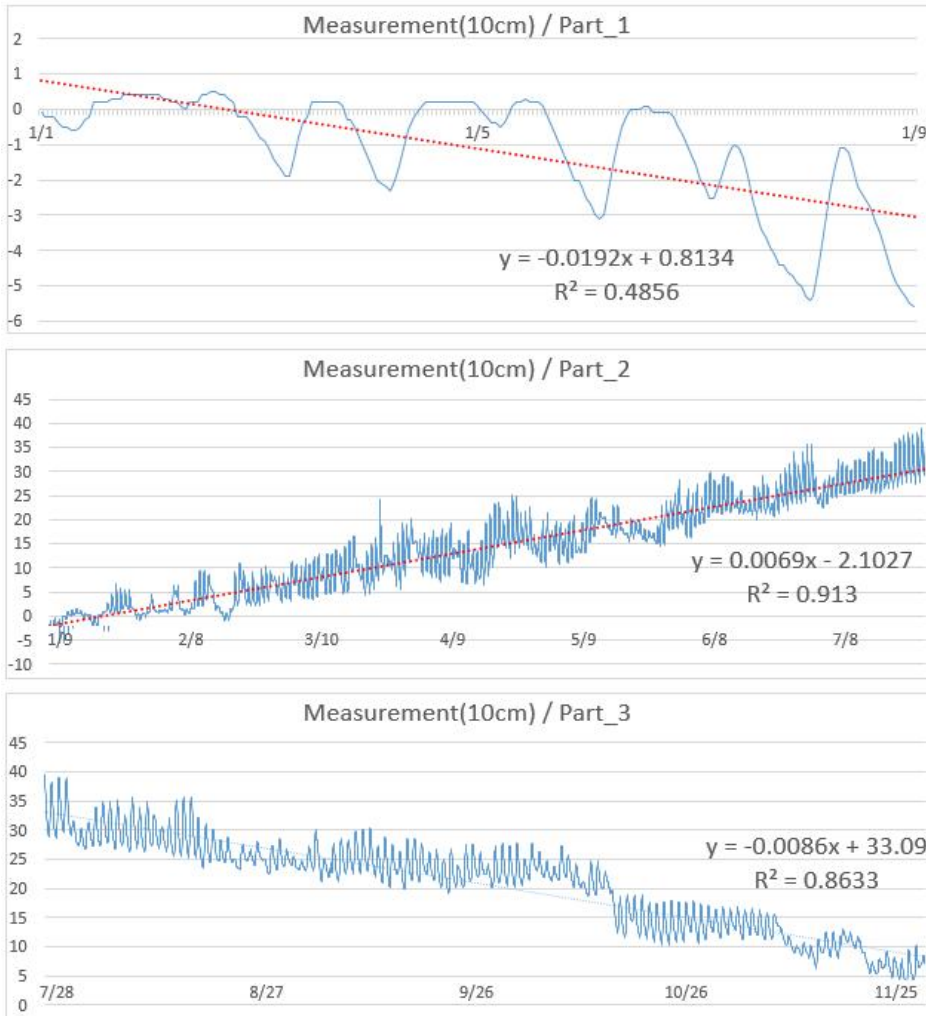


Figure 57 Graph of annual changes in measured values  
at a depth of 10 cm

실측값 10cm 깊이에서의 경우 1월 9일 8시까지 감소하는 추세를 보이며  $-5.6^{\circ}\text{C}$ 로 최저점을 기록 한 이후 7월 28일 17시  $39.6^{\circ}\text{C}$ 로 최고점을 기록할 때까지 상승하는 추세를 보인다. 그리고 그 이후 다시 감소하는 추세를 보인다. 평균값은  $16.26^{\circ}\text{C}$ 이다.

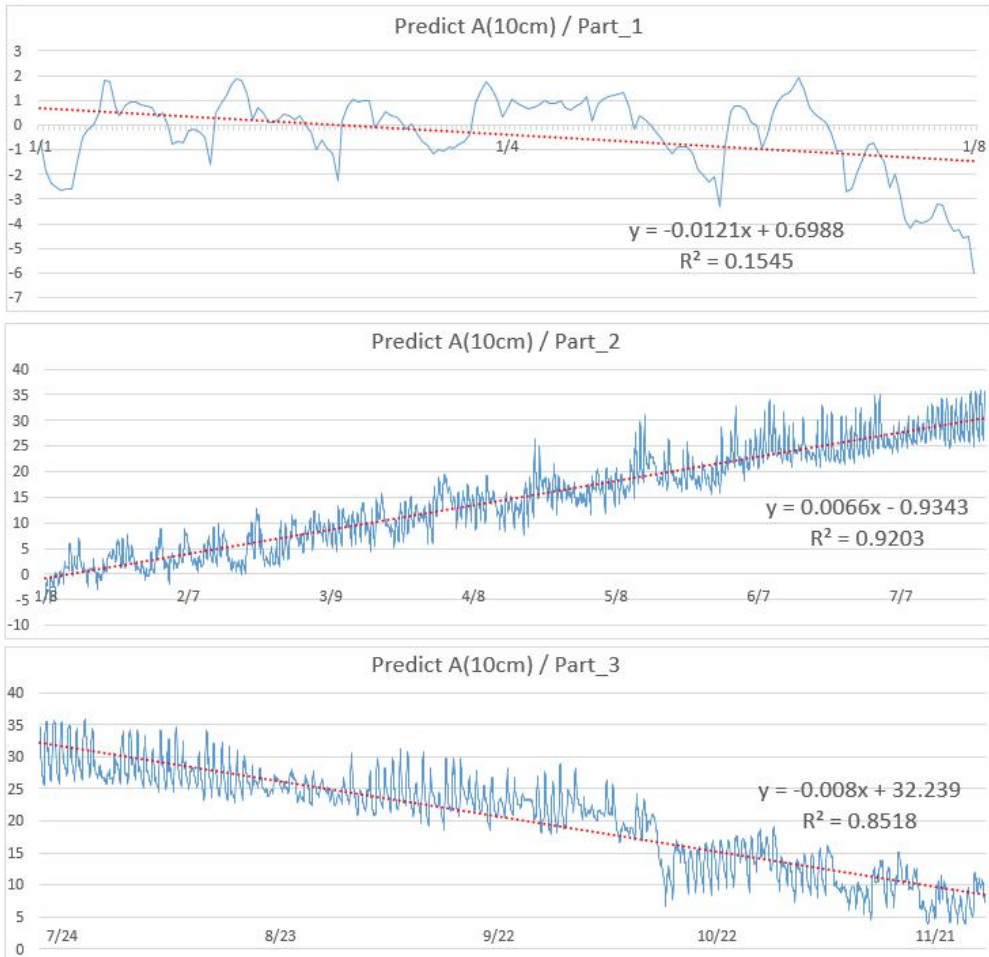


Figure 58 Graph of annual change of predicted value A at a depth of 10 cm.

예측값 A 10cm 깊이에서의 경우 1월 8일 8시  $-5.99^{\circ}\text{C}$ 로 최저점을 기록할 때까지 감소하는 추세를 보이며, 이후 7월 24일 16시  $36^{\circ}\text{C}$ 로 최고점을 기록할 때까지 상승하는 추세를 보인다. 그리고 그 이후 다시 감소하는 추세를 보인다. 평균값은  $16.48^{\circ}\text{C}$ 이다.

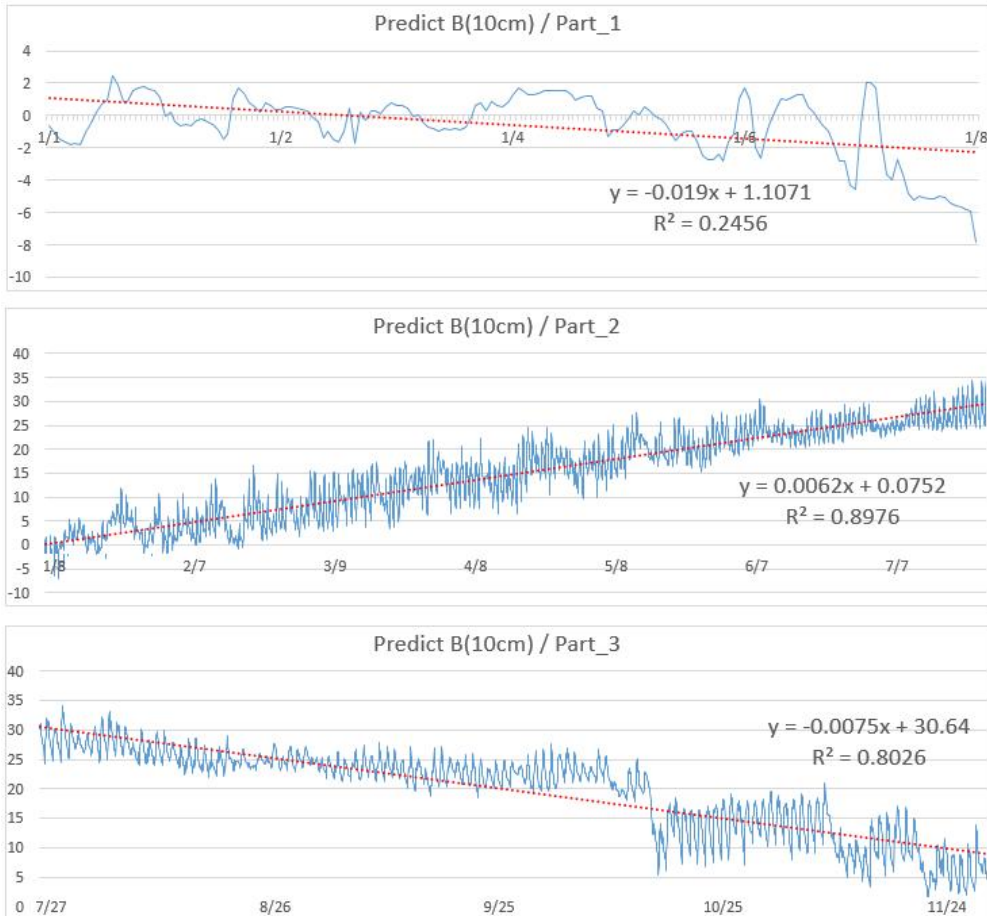


Figure 59 Graph of annual change of predicted value B at a depth of 10 cm.

예측값 B 10cm 깊이에서의 경우 1월 8일 8시  $-7.81^{\circ}\text{C}$ 로 최저점을 기록할 때까지 감소하는 추세를 보이며 이후 7월 27일 15시  $35.3^{\circ}\text{C}$ 로 최고점을 기록할 때까지 상승하는 추세를 보인다. 그리고 그 이후 다시 감소하는 추세를 보인다. 평균값은  $16.28^{\circ}\text{C}$ 이다.

시간과 값의 오차가 다소 발생하였지만 두 예측값 모두 연간 변화량 그래프에서 실측값과 비슷한 추세와 값을 보이는 것을 확인할 수 있다. 세 개의 데이터를 비교한 표는 Table 8 과 같다.

Table 8 Comparison table of annual best records, lowest records, and average values of measured values, predicted values A, and predicted values B

Data	Highest record	Lowest record	Average
Measurement	7/28/17:00, 39.6℃	1/9/8:00, -5.6℃	16.26℃
Predict A	7/24/16:00, 36℃	1/8/8:00, -5.99℃	16.48℃
Predict B	7/27/15:00, 35.3℃	1/8/8:00, -7.81℃	16.28℃

10cm 깊이에서 예측 값 A의 경우 최저점은 실측값과 같은 시간대에 하루 간격의 차이가 있었으며, 0.39℃ 값의 차이가 있었다. 최고점의 경우 일차는 4일, 한 시간 간격의 시간 차이가 있었으며, 3.6℃의 값의 차이가 있었다. 평균값 은 0.22℃ 차이가 있었다. 미세한 차이가 있었지만 예측 값 A는 10cm 깊이에서도 연간 패턴이 동일하게 나타났다고 판단할 수 있다. 예측 값 B의 경우 최저점은 실측값과 같은 시간대에 하루 간격의 차이가 있었으며 2.21℃의 비교적 큰 값의 차이가 있었다. 평균값은 0.02℃ 차이가 있었다. 예측 값 B 역시 10cm 깊이에서는 연간 패턴이 동일하게 나타났다고 판단할 수 있다.

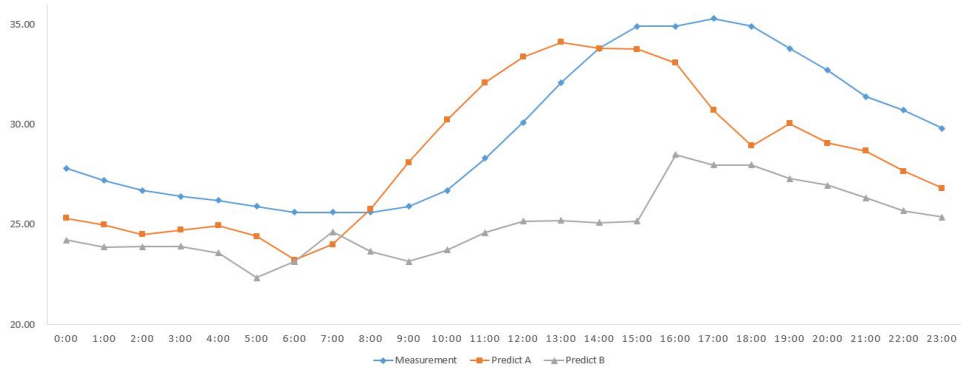


Figure 60 Daily variation graph of the measured value, predicted value A, and predicted value B at August 15 at a depth of 10 cm.

Table 9 Comparison table of the daily highest, lowest, and average values of August 15 of the measured value, predicted value A, and predicted value B

Data	Highest record	Lowest record	Average
Measurement	17:00, 35.30°C	8:00, 25.60°C	29.68°C
Predict A	13:00, 34.11°C	6:00, 23.23°C	28.43°C
Predict B	16:00, 28.49°C	5:00, 22.34°C	25.06°C

Figure 60 과 Table 8. 은 10cm 깊이에서 8월 15일 세 개의 데이터의 변화량을 비교한 그래프인데, 연간 그래프와 다르게 일간 그래프로 자세하게 확인하면 추세의 차이가 있음을 확인할 수 있다. 실측값의 경우 8시에 25.60°C로 최저점을 기록하였고, 예측값 A는 6시에 23.23°C, 예측값 B는 5시에 22.34°C로 최저점을 기록하였다. 최고점은 실측값 17시 35.30°C, 예측값 A 13시 34.11°C, 예측값 B 16시 28.49°C를 기록하였는데 예측값의 값의 오차가 크게 나타나는 것이 확인되었다. 평균값의 경우 실측값 29.68°C, 예측값 A 28.43°C, 예측값 B 25.06°C로 역시 예측값 B가 평균값에서도 비교적 차이가 있었다.

#### 4.3.1.2. 20cm

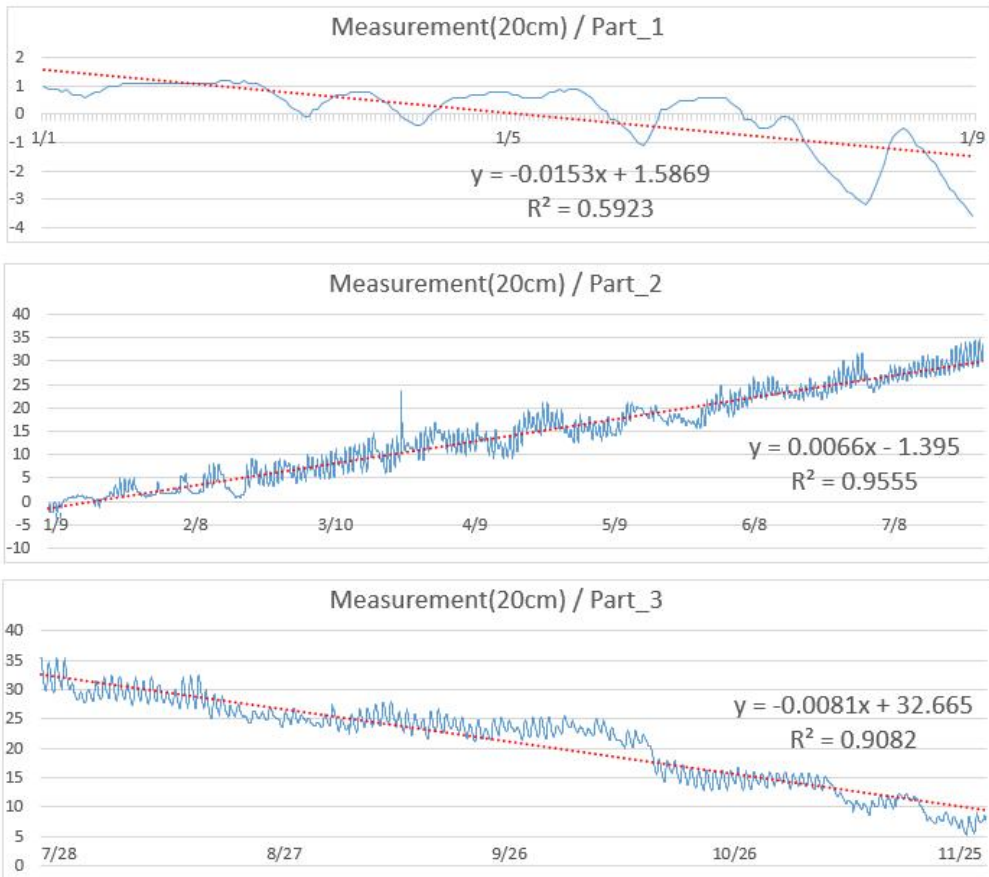


Figure 61 Graph of annual changes in measured values  
at a depth of 20 cm

실측값 20cm 깊이에서의 경우 1월 9일 8시까지 감소하는 추세를 보이며  $-3.6^{\circ}\text{C}$ 로 최저점을 기록 한 이후 7월 28일 17시  $35.3^{\circ}\text{C}$ 로 최고점을 기록할 때까지 상승하는 추세를 보인다. 그리고 그 이후 다시 감소하는 추세를 보인다. 평균값은  $16.42^{\circ}\text{C}$ 이다.

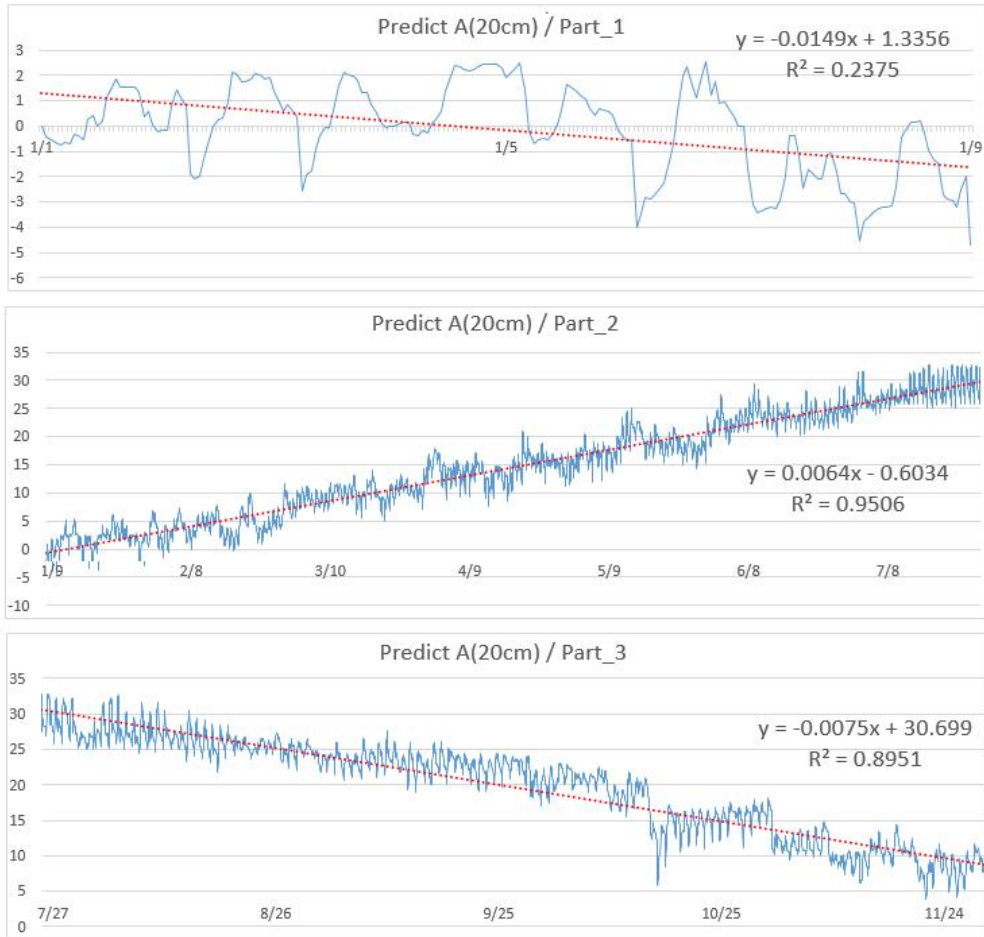


Figure 62 Graph of annual change of predicted value A at a depth of 20 cm.

예측값 A는 20cm 깊이에서의 경우 1월 9일 8시  $-4.7^{\circ}\text{C}$ 로 최저점을 기록할 때까지 감소하는 추세를 보이며, 이후 7월 27일 15시  $32.89^{\circ}\text{C}$ 로 최고점을 기록할 때까지 상승하는 추세를 보인다. 그리고 그 이후 다시 감소하는 추세를 보인다. 평균값은  $16.13^{\circ}\text{C}$ 이다.

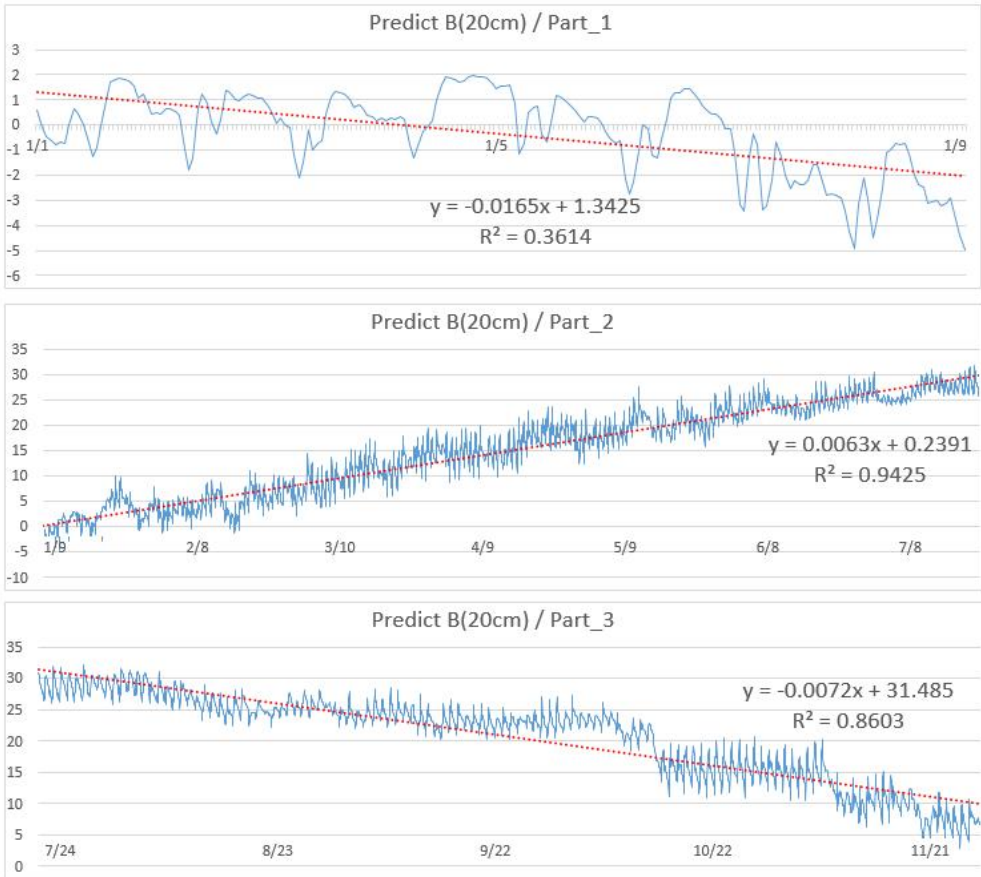


Figure 63 Graph of annual change of predicted value B at a depth of 20 cm.

예측값 B의 20cm 깊이의 경우 1월 9일 10시  $-4.99^{\circ}\text{C}$ 로 최저점을 기록할 때까지 감소하는 추세를 보이며, 이후 7월 24일 16시  $32.11^{\circ}\text{C}$ 로 최고점을 기록할 때까지 상승하는 추세를 보인다. 그리고 그 이후 다시 감소하는 추세를 보인다. 평균값은  $16.78^{\circ}\text{C}$ 이다.

시간과 값의 오차가 다소 발생하였지만 두 예측값 모두 연간 변화량 그래프에서 실측값과 비슷한 추세와 값을 보이는 것을 확인할 수 있다.



세 개의 데이터를 비교한 표는 Table 10과 같다.

Table 10 Comparison table of annual best records, lowest records, and average values of measured values, predicted values A, and predicted values B

Data	Highest record	Lowest record	Average
Measurement	7/28/17:00, 35.3℃	1/9/8:00, -3.6℃	16.42℃
Predict A	7/27/15:00, 32.89℃	1/9/8:00, -4.7℃	16.13℃
Predict B	7/24/16:00, 32.11℃	1/9/10:00, -4.99℃	16.78℃

20cm 깊이에서 예측 값 A의 경우 최저점은 실측값과 같은 시간 같은 날에 발생했으며, 1.1℃값의 차이가 있었다. 최고점의 경우 일자는 하루, 두 시간 간격의 시간 차이가 있었으며, 2.41℃의 값의 차이가 있었다. 평균값 은 0.29℃ 차이가 있었다. 미세한 차이가 있었지만 예측 값 A는 20cm 깊이 에서도 연간 패턴이 동일하게 나타났다고 판단할 수 있다. 예측 값 B의 경우 최저점은 실측값과 같은 한 시간 4일 간격의 차이가 있었으며 3.19℃의 비교적 큰 값의 차이가 있었다. 평균값은 0.36℃ 차이가 있었다. 예측 값 B역시 20cm 깊이 에서 연간 패턴이 동일하게 나타났다고 판단할 수 있다.

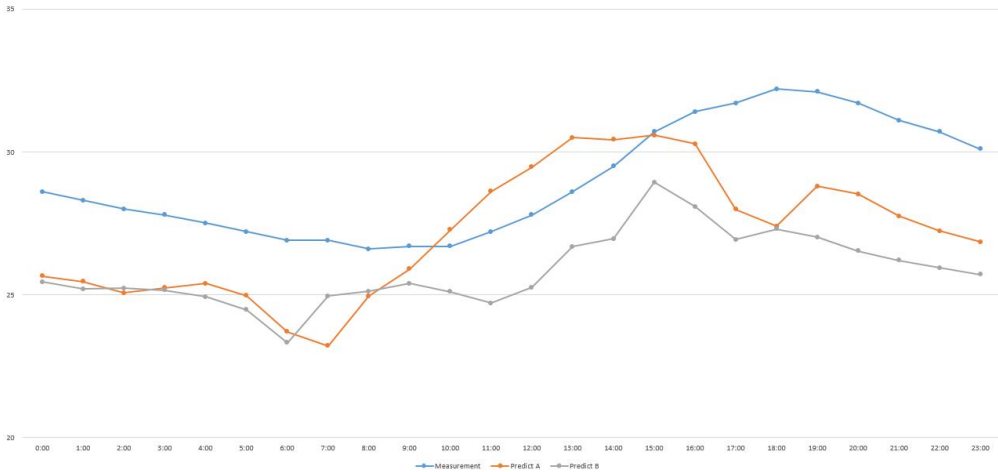


Figure 64 Daily variation graph of the measured value, predicted value A, and predicted value B at August 15 at a depth of 20 cm.

Table 11 Comparison table of the daily highest, lowest, and average values of August 15 of the measured value, predicted value A, and predicted value B

Data	Highest record	Lowest record	Average
Measurement	18:00, 32.2°C	8:00, 26.6°C	29°C
Predict A	15:00, 30.58°C	7:00, 23.21°C	27.14°C
Predict B	15:00, 28.93°C	6:00, 23.32°C	25.86°C

Figure 64 와 Table 11 은 20cm 깊이에서 8월 15일 세 개의 데이터의 변화량을 비교한 그래프와 표인데, 10cm 깊이와 마찬가지로 차이가 보였다. 실측값의 경우 8시에 26.6°C로 최저점을 기록하였고, 예측값 A는 7시에 23.21°C, 예측값 B는 6시에 23.32°C로 최저점을 기록하였다. 최고점은 실측값 18시 32.2°C, 예측값 A와 예측값 B는 15시에 각각 30.58°C과 28.93°C를 기록하였다. 평균값의 경우 실측값 29°C, 예측값 A 27.14°C, 예측값 B 25.86°C로 역시 예측값 B가 평균값에서도 비교적 차이가 있었다.

### 4.3.1.3. 30cm

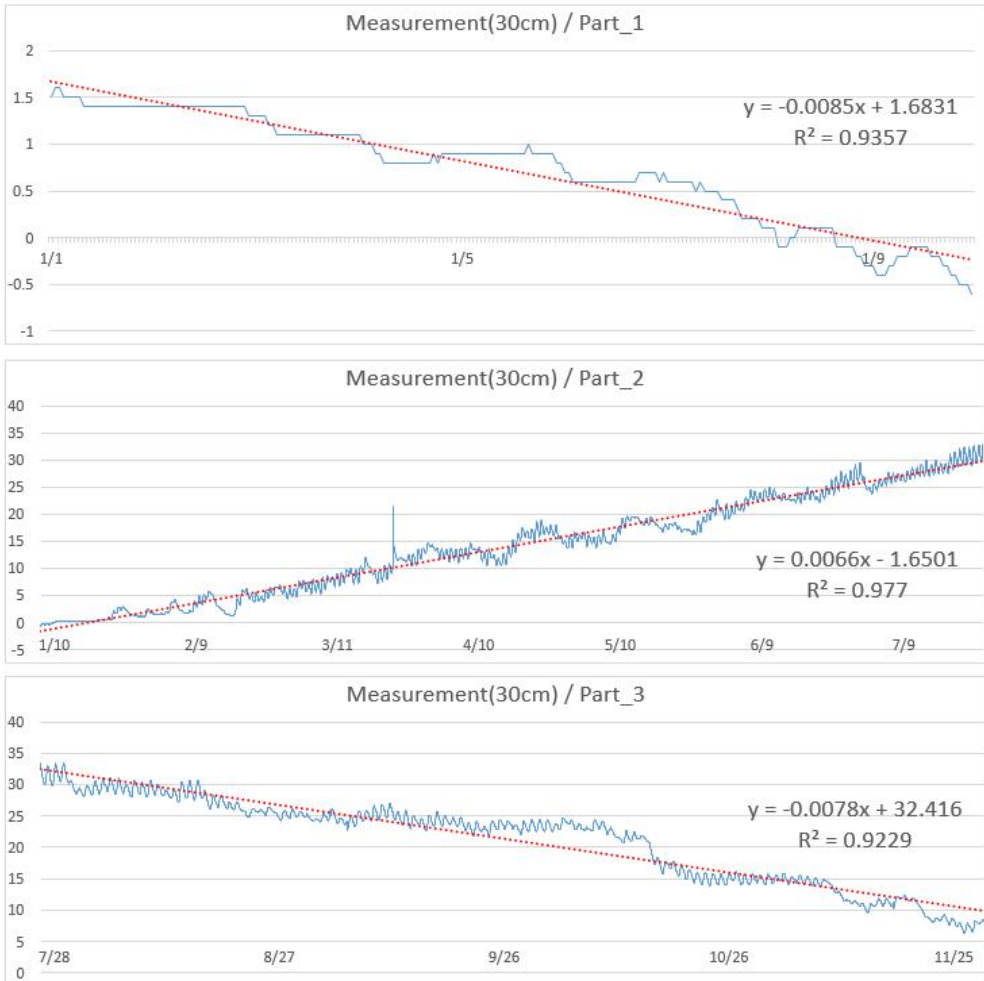


Figure 65 Graph of annual changes in measured values at a depth of 30 cm

실측값 30cm 깊이에서의 경우 1월 10일 8시까지 감소하는 추세를 보이며  $-0.6^{\circ}\text{C}$ 로 최저점을 기록 한 이후 7월 28일 19시  $33.4^{\circ}\text{C}$ 로 최고점을 기록할 때까지 상승하는 추세를 보인다. 그리고 그 이후 다시 감소하는 추세를 보인다. 평균값은  $16.27^{\circ}\text{C}$ 이다.

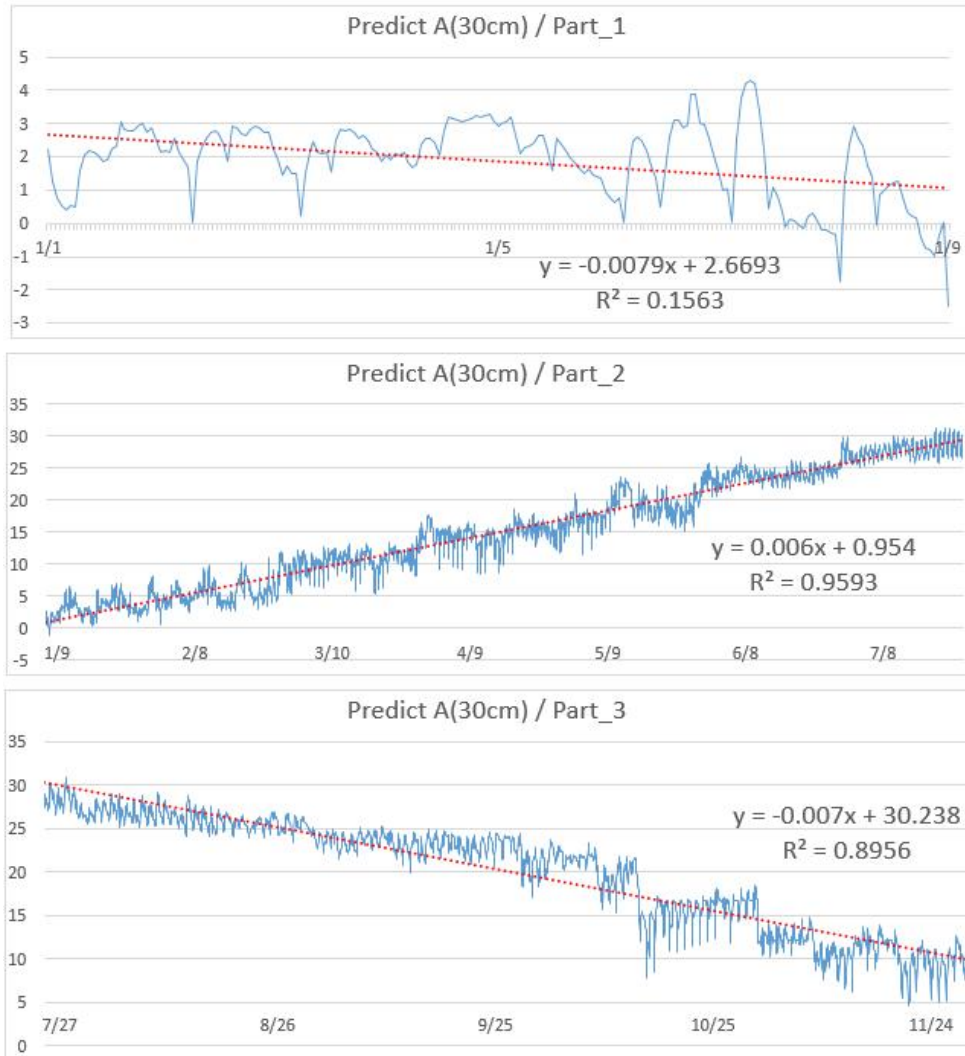


Figure 66 Graph of annual change of predicted value A at a depth of 30 cm.

예측값 A는 30cm 깊이에서의 경우 1월 9일 8시  $-2.5^{\circ}\text{C}$ 로 최저점을 기록할 때까지 감소하는 추세를 보이며, 이후 7월 27일 16시  $31.46^{\circ}\text{C}$ 로 최고점을 기록할 때까지 상승하는 추세를 보인다. 그리고 그 이후 다시 감소하는 추세를 보인다. 평균값은  $16.67^{\circ}\text{C}$ 이다.



Figure 67 Graph of annual change of predicted value B at a depth of 30 cm.

예측값 B의 30cm 깊이의 경우 1월 9일 5시 -5.58℃로 최저점을 기록할 때까지 감소하는 추세를 보이며, 이후 7월 24일 15시 36.3℃로 최고점을 기록할 때까지 상승하는 추세를 보인다. 그리고 그 이후 다시 감소하는 추세를 보인다. 평균값은 16.01℃이다.

최저점의 경우 실측값과 예측값A, B의 오차가 이전 깊이에서 보다 크게 나타나는 것을 확인하였다. 특히 예측값 B의 경우 최저점의 오차가

5.52℃로 매우 크게 나타났다. 세 개의 데이터를 비교한 표는 Table 12와 같다.

Table 12 Comparison table of annual best records, lowest records, and average values of measured values, predicted values A, and predicted values B

Data	Highest record	Lowest record	Average
Measurement	7/28/19:00, 33.4℃	1/10/8:00, -0.6℃	16.27℃
Predict A	7/27/16:00, 31.46℃	1/9/8:00, -2.5℃	16.67℃
Predict B	7/24/15:00, 36.3℃	1/9/5:00, -5.58℃	16.01℃

30cm 깊이에서 예측 값 A의 경우 최저점은 실측값과 같은 시간대에 하루 간격의 차이가 있었으며, 1.9℃ 값의 차이가 있었다. 최고점의 경우 일자는 하루, 3시간 간격의 시간 차이가 있었으며, 1.94℃ 값의 차이가 있었다. 평균값은 0.4℃ 차이가 있었다. 미세한 차이가 있었지만 예측 값 A는 30cm 깊이에서도 연간 패턴이 동일하게 나타났다고 판단할 수 있다. 예측 값 B의 경우 최저점은 실측값과 하루, 3시간 간격의 시간 차이가 있었으며 5.52℃의 비교적 큰 값의 차이가 있었다. 평균값은 0.26℃ 차이가 있었다. 예측 값 B는 30cm 깊이에서도 연간 패턴이 동일하게 나타났다고 판단할 수 있으나, 깊이에 따라 기후가 지온에 미치는 영향력이 감소하여 온도 변화의 폭이 감소하는 패턴은 적용되지 못한 것으로 판단된다.

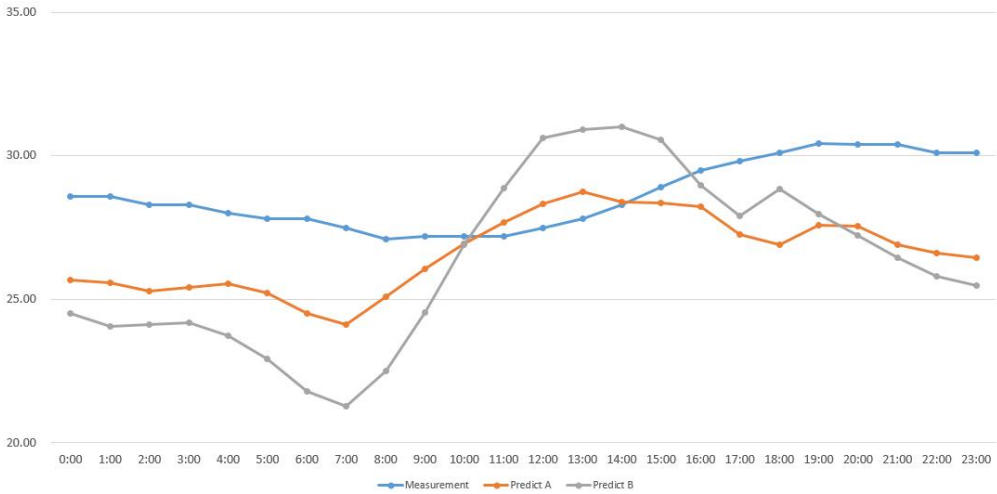


Figure 68 Daily variation graph of the measured value, predicted value A, and predicted value B at August 15 at a depth of 30 cm.

Table 13 Comparison table of the daily highest, lowest, and average values of August 15 of the measured value, predicted value A, and predicted value B

Data	Highest record	Lowest record	Average
Measurement	19:00, 30.42°C	8:00, 27.1°C	28.62°C
Predict A	13:00, 28.75°C	7:00, 24.1°C	26.6°C
Predict B	14:00, 31.03°C	7:00, 21.28°C	26.3°C

Figure 68 과 Table 13 은 30cm 깊이에서 8월 15일 세 개의 데이터의 변화량을 비교한 그래프와 표이다. 예측값 A의 경우 실측값과 최고점의 시간 차이가 6시간이 나지만 실측값과 예측값 A 모두 최고점을 기록한 시간의 값이 근처 시간에서의 값과 차이가 적기 때문에 크게 차이 난다고 볼 수 없다. 실측값에서 나타났던 깊이에 따라 변동 폭이 감소하는 경향이 예측값 A에도 반영되었다. 예측값 B의 경우 실측값과 패턴의 차이가 확연하게 보였다. 깊이에 따라 변동폭이 감소하는 경향이 30cm 깊이에서도 보이지 않았으며, 최고점의 시간 차이가 실측값과 5시간의 차이가 있

었다. 예측값 A처럼 최고점 근처 시간대의 온도가 완만하지 않기 때문에 이는 큰 비교적 큰 차이를 보인다고 판단할 수 있다.

### 4.3.2. 잔차 분석

#### 4.3.2.1. Q-Q Plot을 이용한 정규성 진단

각 예측 값들과 실측값 간에 발생한 잔차의 정규성 검토를 위해 Q-Q Plot(Quantile-Quantile Plot)을 사용하였다. Q-Q Plot은 정규성을 평가하기 위한 일반적인 시각적 기법이며, 수집 데이터를 표준 정규분포의 분위수와 비교하여 그리는 그래프이다. 모집단이 정규성을 보일수록 선형성을 보이며 잔차가 대각선 방향의 직선을 따르면 정규성이 있다고 해석하며, 잔차가 정규성일 보일수록 회귀분석이 잘 되었다고 평가할 수 있다.

Q-Q Plot은 수치로 평가하는 정량적 분석(Quantitative Analysis)이 아닌 시각적으로 평가하는 정성적 분석(Qualitative Analysis)이기 때문에 시각적인 그래프를 통해 정규성을 확인 하였으며 Python 프로그램을 이용해 그래프를 그렸다. 예측값 A와 예측값 B의 각 깊이별 실측값과의 Q-Q Plot 그래프는 Figure 69 와 같다.



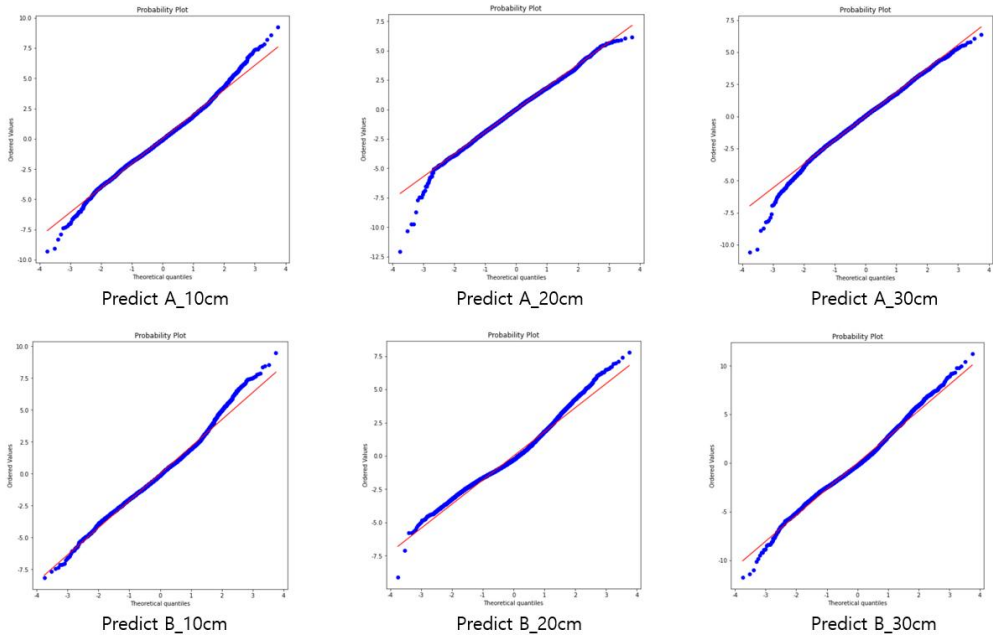


Figure 69 Q-Q plot for each depth of predicted values A and B

예측값 A와 B 모두 모든 깊이에서 각 요소가 대각선 방향의 직선을 따르고 있어 정규성을 따르고 있다고 평가할 수 있다. 다만 평가 지표상으로 비교적 더 좋은 평가를 받았던 예측값 A 보다 예측값 B가 모든 깊이에서 정규분포를 따르는 것을 그래프 통해 평가할 수 있다.

#### 4.3.2.2. 표준화 잔차를 이용한 이상치 검출

본 연구에서는 마지막으로 표준화 잔차(Standardized Residuals)를 이용하여 각 예측값에서 깊이별 이상치(Outliers)를 검출하였고, 이상치들을 가지는 시간대들의 공통점을 확인하고 원인을 유추하는 실험을 하였다.

Micro soft사에서 제공하는 Excel 2013을 활용하였으며, 각 데이터 별로 잔차와 레버리지(leverage)를 통해 확인된 표준화 잔차를 이용해 이상치를 검출하였다. 표준화 잔차의 절대값이 3을 초과한 경우 이상치라고 판단하였고, 양수와 음수를 구분하여 비교하였다.

### 4.3.2.2.1. 예측값 A

Predict A											
10cm				20cm				30cm			
(+)		(-)		(+)		(-)		(+)		(-)	
Time	Value	Time	Value	Time	Value	Time	Value	Time	Value	Time	Value
03-25 15:00	4.65	02-01 10:00	-3.08	03-25 15:00	5.46	05-12 13:00	-3.02	03-25 15:00	4.76	02-20 14:00	-3.24
03-25 16:00	4.48	02-01 11:00	-3.06	03-25 16:00	6.34	05-12 14:00	-3.13	03-25 16:00	5.68	05-12 14:00	-3.19
04-26 16:00	3.54	02-19 21:00	-3.01	04-26 17:00	3.07	05-14 12:00	-3.03	10-11 09:00	3.02	06-04 10:00	-3.27
04-26 17:00	3.63	02-19 22:00	-3.02	07-27 18:00	3.32			10-11 13:00	3.04		
04-26 18:00	3.12	02-19 23:00	-3.01	07-28 18:00	3.02			10-11 14:00	3.45		
06-28 18:00	3.36	02-21 10:00	-3.64	07-29 17:00	3.16			10-11 15:00	3.25		
06-30 18:00	3.18	02-21 11:00	-3.15	07-29 18:00	3.16			10-16 09:00	3.38		
07-01 18:00	3.30	05-12 11:00	-3.54	07-31 18:00	3.63			10-16 10:00	3.51		
07-27 18:00	4.27	05-12 13:00	-3.31	10-16 10:00	3.09			10-16 11:00	4.10		
07-28 17:00	3.82	05-12 14:00	-3.17	10-16 11:00	3.64			10-16 12:00	4.37		
07-28 18:00	4.02	05-13 09:00	-3.29	10-16 12:00	3.91			10-16 13:00	4.68		
07-29 16:00	3.41	05-13 10:00	-3.57	10-16 13:00	4.04			10-16 14:00	3.76		
07-29 17:00	3.97	05-13 11:00	-3.49	10-16 14:00	3.91			10-16 15:00	3.15		
07-29 18:00	3.62	05-13 12:00	-3.57	10-16 15:00	3.60			10-16 16:00	3.05		
07-31 17:00	3.88	05-13 13:00	-3.20	10-16 16:00	3.33			10-17 07:00	5.48		
07-31 18:00	4.47	05-14 10:00	-3.38	10-17 07:00	4.98			10-17 08:00	4.35		
07-31 19:00	3.23	05-14 11:00	-4.03	10-17 08:00	4.97			10-17 09:00	3.52		
08-06 16:00	3.32	05-14 12:00	-4.33	10-17 09:00	4.44			10-17 10:00	3.27		
08-16 16:00	3.32	05-14 13:00	-3.80	10-17 10:00	3.79			10-17 11:00	3.21		
08-16 17:00	3.51	05-14 14:00	-3.38	10-17 11:00	3.32			10-17 12:00	3.06		
08-16 18:00	3.08	05-23 11:00	-3.74	10-17 12:00	3.17			10-17 14:00	3.11		
08-17 18:00	3.50	06-02 10:00	-3.42	10-17 13:00	3.07			10-17 15:00	3.63		
08-18 17:00	3.53	06-02 11:00	-3.52	10-17 14:00	3.37			10-17 16:00	4.17		
10-16 14:00	3.21	06-07 14:00	-3.12	10-17 15:00	3.63			10-18 07:00	3.61		
10-16 15:00	3.12	06-09 11:00	-3.02	10-17 16:00	3.73			10-18 08:00	4.29		
10-17 07:00	3.05	06-10 12:00	-3.19					10-18 09:00	3.17		
10-17 14:00	3.08	06-20 14:00	-3.04					11-08 13:00	3.01		
10-17 15:00	3.43	06-22 11:00	-3.36					11-08 14:00	3.28		
10-17 16:00	3.61	06-22 12:00	-3.18					11-09 13:00	3.06		
10-17 17:00	3.10							11-22 12:00	3.37		

Figure 70 Time zone and standardized residuals value of outliers by depth of predicted value A.

10cm 깊이에서 이상치의 개수는 양수 30개, 음수 29 총 59개로 총 7867개의 데이터의 0.75%를 차지하였다. 양수의 경우 30개의 이상치 중 28개의 시간대가 14시부터 18시까지 분포해 있었으며, 이 시간대는 태양 에너지의 영향을 받는 시간이다. 10월 17일에 이상치가 많이 분포되어 있는 것으로 확인되었는데, 해당 지역에서의 해당 날짜 기후 특이점은 Figure 71 과 같이 전후 일주일 동안 일 평균 기온과 일 평균 습도가 유독 낮은 것으로 확인되었으며 그 외 특이점은 확인되지 않았다.

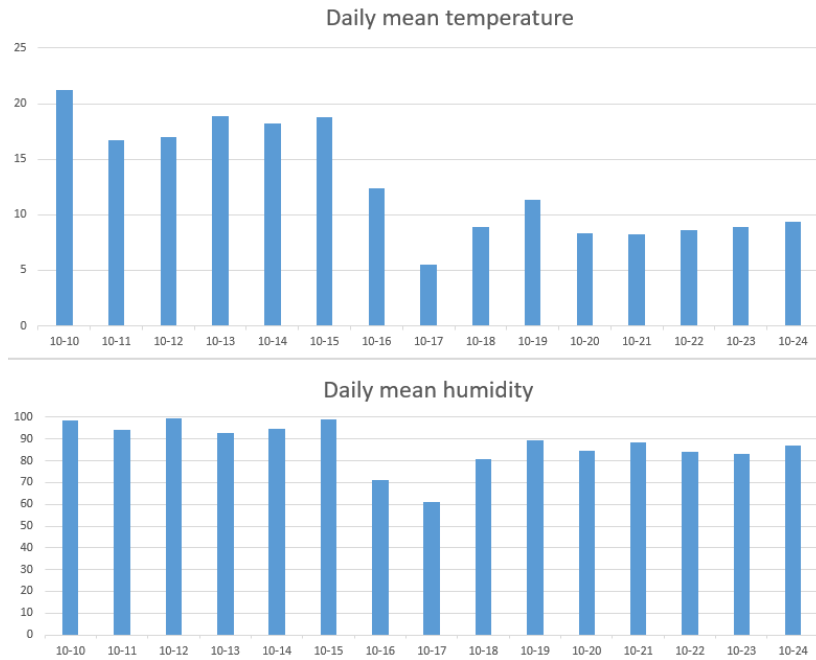


Figure 71 Time zone and standardized residuals value of outliers by depth of predicted value A.

음수 이상치의 경우 29개의 이상치 중 26개가 10시부터 14시까지 분포해 있었으며, 이 시간대 역시 태양 에너지의 영향을 받는 시간대이다. 5월 12일~5월 14일에 유독 이상치가 많이 발생하는 것으로 확인되었고, 해당 지역의 날짜 구간의 기후 특이점은 전후로 강우가 있었다는 것으로 확인되었다.

20cm 깊이에서 이상치의 개수는 양수 25개, 음수 3 총 28개로 총 7867개의 데이터의 0.36%를 차지하였다. 양수, 음수 총 28개의 이상치 중 27개의 시간대가 8시부터 18시까지 분포해 있었으며, 이 시간대 역시 일사량의 영향을 받는 시간이다. 10월 16일과 10월 17일에 많은 이상치가 분포되어 있었으며, 10월 16일의 해당 지역 기후 특이점은 발견하지 못하였다.

30cm 깊이에서 이상치의 개수는 양수 30개, 음수 3 총 33개로 총 7867개의 데이터의 0.41%를 차지하였다.

양수, 음수 총 33개의 이상치 중 31개의 시간대가 8시부터 16시까지 분포해 있었으며, 마찬가지로 일사량의 영향을 받는 시간이다. 30cm 깊이에서도 10월 16일과 10월 17일에 이상치가 많이 분포되어 있음을 확인하였다.

표준화 잔차를 이용한 예측값A의 이상치 검출 결과 모든 깊이에서 전체데이터의 1%미만 으로 높은 성능을 보였다. 이상치의 대부분은 태양에너지의 영향을 받는 시간대에 분포하였다. 모든 깊이에서 10월16일, 10월 17일에 이상치가 많이 분포 되어있다는 것을 확인할 수 있는데 10월17일의 경우 전후 날짜에 비해 일평균 기온과 일평균 습도가 낮게 나타나는 것을 확인할 수 있다.

### 4.3.2.2.2. 예측값B

Predict B															
10cm						20cm						30cm			
(+)		(-)				(+)		(-)				(+)		(-)	
Time	Value	Time	Value	Time	Value	Time	Value	Time	Value	Time	Value	Time	Value	Time	Value
07-02 15:00	3.03	01-24 11:00	-3.54	03-30 07:00	-3.37	03-25 15:00	3.97	01-24 10:00	-3.24	05-01 10:00	-3.17	10-17 00:00	3.02	01-23 15:00	-3.05
07-27 17:00	3.23	01-24 12:00	-3.63	03-30 08:00	-3.82	03-25 16:00	5.03	01-24 11:00	-3.70	05-01 11:00	-3.43	10-17 01:00	3.07	01-23 16:00	-3.08
07-28 16:00	3.32	01-25 11:00	-3.43	03-30 09:00	-3.24	08-17 17:00	3.36	01-24 12:00	-3.64	05-08 07:00	-3.36	10-17 02:00	3.11	01-23 17:00	-3.11
07-28 17:00	3.30	02-01 11:00	-3.24	03-31 08:00	-3.12	08-17 18:00	3.30	01-24 13:00	-3.24	05-11 08:00	-3.41	10-17 03:00	3.25	01-23 18:00	-3.04
07-29 17:00	3.06	02-01 13:00	-3.08	03-31 09:00	-3.35	08-17 19:00	3.44	01-25 11:00	-3.62	05-11 14:00	-3.23	10-17 04:00	3.34	01-25 14:00	-3.38
07-31 16:00	3.58	02-05 13:00	-3.09	03-31 10:00	-3.03	08-17 20:00	3.43	01-25 12:00	-3.47	05-14 14:00	-3.64	10-17 05:00	3.44	01-25 15:00	-3.49
07-31 17:00	3.53	02-12 11:00	-3.27	03-31 11:00	-3.93	08-17 21:00	3.20	01-25 13:00	-3.08	05-14 15:00	-3.39	10-17 06:00	4.23	01-25 16:00	-3.09
08-15 14:00	3.36	02-12 12:00	-3.05	03-31 12:00	-3.36	08-17 22:00	3.12	01-25 14:00	-3.06	05-22 15:00	-3.34	10-17 07:00	4.11	02-21 13:00	-3.24
08-15 15:00	3.79	02-13 11:00	-4.65	04-01 09:00	-3.12			03-05 09:00	-3.07	05-23 12:00	-3.76	10-17 08:00	3.95	02-21 14:00	-3.52
08-16 13:00	3.06	02-13 12:00	-3.86	04-01 12:00	-3.07			03-05 10:00	-3.35	05-23 13:00	-3.93	10-17 09:00	3.63	02-21 15:00	-3.50
08-16 16:00	3.07	02-19 11:00	-3.34	04-17 07:00	-3.18			03-05 11:00	-3.10	05-25 15:00	-3.03	10-17 10:00	3.31	02-21 16:00	-3.13
08-17 14:00	3.63	02-19 12:00	-3.82	04-19 07:00	-3.11			03-17 09:00	-3.21	06-07 15:00	-3.13	10-18 01:00	3.00	04-21 12:00	-3.05
08-17 15:00	3.83	02-19 13:00	-3.75	04-19 08:00	-3.18			03-18 09:00	-3.51	10-29 13:00	-3.31	10-18 02:00	3.02	04-21 13:00	-3.25
08-17 16:00	3.12	02-20 11:00	-3.84	04-20 07:00	-3.20			03-25 08:00	-3.01	11-06 11:00	-3.11	10-18 03:00	3.02	04-21 14:00	-3.37
08-17 17:00	3.47	02-20 12:00	-3.68	04-21 09:00	-3.31			03-30 08:00	-3.39	11-06 12:00	-3.13	10-18 05:00	3.17	04-21 15:00	-3.22
08-17 18:00	3.22	02-21 09:00	-3.09	04-21 10:00	-3.14			03-30 09:00	-3.05	11-07 10:00	-3.21	10-18 06:00	3.50	05-13 13:00	-3.09
08-17 19:00	3.05	02-21 10:00	-3.51	04-22 09:00	-3.19			04-01 07:00	-3.05	11-07 11:00	-3.53	10-18 07:00	3.14	05-13 14:00	-3.22
08-18 16:00	3.14	02-21 11:00	-3.63	05-12 08:00	-3.22			04-09 07:00	-3.36			05-14 12:00	3.43	05-14 12:00	-3.42
10-16 12:00	3.05	02-21 12:00	-3.61	05-22 15:00	-3.11			04-14 07:00	-3.02			05-14 13:00	3.94	05-14 13:00	-3.94
10-16 13:00	3.32	02-21 13:00	-3.55	05-23 12:00	-3.33			04-15 08:00	-3.00			05-14 14:00	3.63	05-14 14:00	-3.63
10-16 14:00	3.32	02-26 12:00	-3.67	05-23 13:00	-3.38			04-17 07:00	-3.75			05-14 15:00	3.45	05-14 15:00	-3.45
10-16 15:00	3.16	02-26 13:00	-3.18	05-26 11:00	-3.40			04-17 08:00	-4.33			06-09 12:00	3.29	06-09 12:00	-3.29
10-17 06:00	3.63	03-05 09:00	-4.20	05-26 12:00	-3.27			04-17 12:00	-3.43			06-09 13:00	3.10	06-09 13:00	-3.10
10-17 15:00	3.54	03-05 10:00	-4.13	05-28 12:00	-3.07			04-17 13:00	-3.67			06-09 14:00	3.06	06-09 14:00	-3.06
10-17 16:00	3.25	03-05 11:00	-3.72	05-28 13:00	-3.03			04-17 14:00	-3.01						
		03-06 11:00	-3.29	05-28 14:00	-3.16			04-17 15:00	-3.46						
		03-08 11:00	-3.63	11-07 12:00	-3.21			04-19 07:00	-3.86						
		03-08 12:00	-3.02	11-15 09:00	-3.26			04-19 08:00	-3.23						
		03-10 10:00	-3.31	11-29 11:00	-3.50			04-19 15:00	-3.19						
		03-10 11:00	-3.13					04-20 07:00	-3.63						
		03-15 10:00	-3.03					04-20 08:00	-3.22						
		03-23 09:00	-3.36					04-21 07:00	-3.37						
		03-23 10:00	-3.32					04-25 07:00	-3.58						
								04-30 12:00	-3.52						

Figure 72 Time zone and standardized residuals value of outliers by depth of predicted value B.

10cm 깊이에서 이상치의 개수는 양수 25개, 음수 62 총 87개로 총 7867개의 데이터의 1.11%를 차지하였다. 양수의 경우 25개의 이상치 중 23개의 시간대가 12시부터 18시까지 분포해 있었으며, 이 시간대는 태양 에너지의 영향을 받는 시간이다. 8월 15일~8월17일 동안 이상치가 많이 발생하는 것으로 확인 되었는데, 해당 지역에서의 이 기간 동안의 기후 특이점은 8월15일 전후 날짜에 비해 습도가 낮은 경향을 보이는 것을 제외하고는 발견되지 않았다. 예측값 A에서 많은 이상치를 보였던 10월16일과 10월17에도 이상치가 분포하는 것을 확인할 수 있다. 음수의 경우 62개의 이상치 중 58개의 시간대에서 8시부터 14시까지 분포해 있었으며, 이 시간대 역시 태양 에너지의 영향을 받는 시간이다. 2월과 3월에 이상치가 많이 분포되어 있는 것을 확인할 수 있다.

20cm 깊이에서 이상치의 개수는 양수 8개, 음수 51 총 59개로 총 7867개의 데이터의 0.75%를 차지하였다. 대부분의 이상치가 음수로 발생되었고 경우 총 59개의 이상치 중 46개의 시간대가 8시부터 18시까지 분포되어있다. 이 시간대는 태양 에너지의 영향을 받는 시간이다. 3월, 4월, 5월에 주로 분포되어 있다.

30cm 깊이에서는 이상치의 개수는 양수 17개, 음수 24 총 41개로 총 7867개의 데이터의 0.52%를 차지하였다. 시간상의 특징은 보이지 않았으며, 앞선 결과들에서 많은 이상치가 분포되어있던 10월17일과 그 다음날 10월18일에 모두 분포되어있는 것으로 확인되었다. 음수의 경우 24개의 이상치 모두 12시부터 18시까지의 시간대에 분포해 있었으며, 양 에너지의 영향을 받는 시간이다.

## 5. 결론

본 연구에서는 다섯 개의 기상인자와 세가지 인공지능 모델 Recurrent neural network(RNN), Long Short-Term Memory(LSTM), Gatedrecurrent unit(GRU)을 활용하여 깊이별 지온을 예측하고 예측값의 정확성을 통해 실제 노지 농업에서의 적용여부를 평가하였다.

기상인자 데이터와 깊이별 지온데이터를 통해 예측모델을 만들고 동일 지역에서 그대로 적용 시켜 성능을 평가하던 기존의 연구들과 달리 본 연구에서는 비용과 시간을 절약하기 위해 무료로 제공되는 공공의 기상인자 데이터와 깊이별 지온데이터를 통해 예측 모델을 만들고 실제 농업활동이 이루어지고 있는 특정 밭 지역에서 예측모델을 사용하여 성능을 평가하였다.

기온, 습도, 풍향, 풍속, 일사량 다섯 가지 기상인자의 시간단위 데이터와 10cm, 20cm, 30cm 세 곳의 깊이의 시간단위 데이터를 활용하였고 실험은크게 세 가지로 나뉘어 진행하였다.

먼저 기상인자 데이터를 이용하여 최적의 매개변수를 선택하기 위한 실험을 진행하였다. 각 기상인자를하나씩 소거하여 생성된 각각의 예측값들을 모든 기상인자를 사용하여 생성한 예측값과 정확도의 증감과 증감의 크기를 확인하여 각 기상인자가 깊이별 지온을 예측하는데 미치는 영향력을 확인하였다. 그리고 전진선택법을 통해 영향력의 크기 순서대로 투입하여 최적의 데이터셋을 구성하는 실험을 하였다.

두 번째 실험은 주기성을 가지고 있는 인자를 매개변수로 투입하였을 시 연간 주기와 주행성에 영향을 받는 지온을 예측하는데 있어 정확도 상승에 영향을 미치는지 여부를 확인하는 실험을 하였다. 전국 월평균기온, 태양 고도, 타 지역 깊이별 지온값 을 주기성 매개변수로 사용하였고, 정확도 평가를 통해 최적의 기상인자 매개변수, 주기성 인자 매개변수, 인공지능 모델 조합을 찾는 최적화 실험을 하였다.

그리고 모든 깊이에서 높은 정확도를 가지고 있다고 판단되는 데이터세트, 인공지능 모델 조합 두 개를 선택하였다.

마지막 실험은 전 실험들의 결과에 따라 성능이 가장 높게 평가된 두 개의 데이터세트, 모델 조합의 예측값 들을 실측값과 비교하여 실제 농업 활동에 적용 가능한지 평가하였다. 연간, 일간 패턴을 비교하여 유사성을 확인하였고, 잔차분석을 통해 정규성을 진단하고 이상치를 검출하는 실험을 진행하였다.

각 단계의 과정과 결과를 정리하면 다음과 같다.

- 1) 다섯 개의 기상인자 매개변수로 활용하여 인공지능 모델을 통해 깊이 별 지온 예측을 하였을 때, 변수의 영향력은 기온, 습도, 일사, 풍속, 풍향 순서로 크게 나타났고, 기온의 경우 다른 기상인자에 비해 월등히 영향력이 크게 나타났지만 깊이가 증가함에 따라 영향력이 점점 감소하였다. 이는 지표면에서 교환된 열 에너지가 깊이에 따라 감소되기 때문이라고 유추할 수 있다.

풍향의 경우 예측을 하는데 있어 정확도를 떨어뜨리는 요인이 되었다. 따라서 이후의 실험에서 풍향은 매개변수로써 제외 하였다.

기온, 습도, 일사, 풍속 매개변수를 순서대로 하나씩 투입하여 실험을 진행한 결과 매개변수가 추가될수록 작은 폭으로 정확도가 높은 예측을 하였고, 깊이가 깊어질수록 이러한 양상은 더 확연히 나타났다.

- 2) 주기를 나타내는 매개변수로써 전국 월평균기온, 태양고도, 타지역 깊이 별 지온 데이터를 각각 활용하여 실험을 진행 했을 시 태양고도를 주기성 매개변수로 활용하였을 시 정확도가 감소하는 것을 확인하였다. 월평균기온과, 타지역 깊이 별 지온데이터를 주기성 매개변수로 활용 하였을 시 전체적으로 정확도가 상승하였지만 반대의 양상을 보이는 경우도 있었다.

특히 주기성 매개변수를 활용 하였을 시에는 모델에 따른 정확도의 차이가 기상인자 매개변수만 활용 하였을 때 보다 크게 나타났다.

평가지표를 통해 모든 깊이에서 전반적으로 가장 좋은 성능을 보였던 데이터 세트와 인공지능 모델 조합 두 개는 다음과 같다고 평가 하였



다.

A. 기온, 습도, 풍속, 일사량, 월 평균기온 + LSTM 모델

B. 기온, 습도, 풍속, 일사량, 타 지역 깊이별 지온데이터 + RNN 모델

- 3) 이전 실험에서 가장 성능이 좋았던 데이터 세트와 인공지능 모델 조합 예측값을 각각 예측값A, 예측값B로 지정하고 실측값과 패턴분석과, 잔차분석을 통해 실제 노지 농업에 적용가능성을 검토하였다.

두 예측값 모두 실측값과 연간 비슷한 패턴을 보였다. 최고지점의 시간과 값, 최저지점의 시간과 값, 평균값 모두 대부분의 깊이에서 높은 일치성을 보였다. 두 예측값 중 예측값A가 실측값과 더 일치율이 높았다.

Q-Q Plot을 이용하여 정규성을 진단한 결과 두 예측값 모두 잔차의 정규성이 있다고 그래프를 통해 확인 하였으며, 두 예측값 중 B가 더 잔차의 정규성이 있다고 판단 되었다.

표준화 잔차를 이용해 각각 예측값의 깊이 별 이상치를 검출하였다.

예측값 A는 10cm 깊이에서 전체 예측값의 0.75%, 20cm 깊이에서 0.36%, 30cm 깊이에서 0.41%가 발생하여 모든 깊이에서 1%미만의 미량의 이상치가 검출되었다.

예측값 B는 10cm 깊이에서 전체 예측값의 1.11%, 20cm 깊이에서 0.75%, 30cm 깊이에서 0.52%가 발생하여 예측값A에 비해 많은 이상치가 검출 되었지만 낮은 수치가고 판단된다.

- 4) 실험과정을 통해 선별된 두 데이터세트, 모델 조합의 예측값 모두 평가지표, 패턴분석, 잔차분석을 통해 실측값과 높은 일치성을 보인다고 판단되며, 실제 농업활동에 적용 가능하다고 평가하였다.

## 6. 참고문헌

Alizamir, Meysam, et al. "Advanced machine learning model for better prediction accuracy of soil temperature at different depths." PLoS One 15.4 (2020): e0231055.

Bassirirad, Hormoz. "Kinetics of nutrient uptake by roots: responses to global change." The New Phytologist 147.1 (2000): 155-169.

Bilgili, Mehmet. "Prediction of soil temperature using regression and artificial neural network models." Meteorology and atmospheric physics 110.1 (2010): 59-70.

Chen, Ke-Tao, et al. "Design of monitoring system for multilayer soil temperature and moisture based on WSN." 2014 International Conference on Wireless Communication and Sensor Network. IEEE, 2014.

Clarke, Simon J., et al. "Spring root zone temperature regulates root growth, nutrient uptake and shoot growth dynamics in grapevines." Australian journal of grape and wine research 21.3 (2015): 479-489.

Colah's Blog, 2015. Understanding LSTM Networks.

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Deardorff, James W. "Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation." Journal of Geophysical Research: Oceans 83.C4 (1978): 1889-1903.

Dingman SL. 1994. *Physical Hydrology*. Prentice-Hall: Upper Saddle River, NJ.

Enrique, Gonzalez-sosa, et al. "Modelling heat and water exchanges of fallow land covered with plant-residue mulch." *Agricultural and Forest Meteorology* 97.3 (1999): 151-169.

Feng, Yu, et al. "Estimation of soil temperature from meteorological data using different machine learning models." *Geoderma* 338 (2019): 67-77.

Goap, Amarendra, et al. "An IoT based smart irrigation management system using Machine learning and open source technologies." *Computers and electronics in agriculture* 155 (2018): 41-49.

Haggard, Beatrix, et al. "Spatial Soil Temperature and Moisture Monitoring Across the Transylvanian Plain in Romania." *Bulletin of the University of Agricultural Sciences & Veterinary Medicine Cluj-Napoca. Agriculture* 67.1 (2010).

Jo, Gwang-Hui, et al. "Establishment of data for integrated artificial intelligence learning of precision agricultural field crops. (Jeollabuk do)." *Magazine of the Korean Society of Agricultural Engineers* 63.3 (2021): 10-16. (in Korean)

Jungqvist, Gunnar, et al. "Effect of climate change on soil temperature in Swedish boreal forests." *PloS one* 9.4 (2014): e93957.

Kang, Sinkyu, et al. "Predicting spatial and temporal patterns of soil temperature based on topography, surface cover and air temperature." *Forest Ecology and Management* 136.1-3 (2000): 173-184.

Kaspar, T. C., and WI L. Bland. "Soil temperature and root growth." *Soil Science* 154.4 (1992): 290-299.

Minyoung Kim, et al. "Soil Moisture Based-smart Irrigation System for Cucumber Cultivation." *Journal of the Korean Society for Agricultural Machinery*, 23.2 (2018): 226-226.

Lahti, M., et al. "Soil temperature, gas exchange and nitrogen status of 5-year-old Norway spruce seedlings." *Tree physiology* 22.18 (2002): 1311-1316.

Lee, Jae-Hyung. "Study on Winegrape Growth Environment Management Method using Ubiquitous Sensor Network." *Proceedings of the Korea Information Processing Society Conference*. Korea Information Processing Society, 2011.

Lee Wooseop, and Kim Hyungkyoo. "Prediction Model of Average Temperature based on Characteristic of Urban-space Using LSTM and GR U: The Case of Wonju City" *The Korea Spatial Planning Review* (2021): 89-104.

Lewis, Roland W., Perumal Nithiarasu, and Kankanhalli N. Seetharamu. *Fundamentals of the finite element method for heat and fluid flow*. John Wiley & Sons, 2004.

RDA, *Agricultural technology guide 28\_Sweet potatoes*, 2018 [in. Korean]

RDA, *Agricultural technology guide 35\_corn*, 2018 [in. Korean]

RDA, Agricultural technology guide 96\_Onion, 2018 [in. Korean]

Repo, Tapani, et al. "The effect of soil temperature on the bud phenology, chlorophyll fluorescence, carbohydrate content and cold hardiness of Norway spruce seedlings." *Physiologia Plantarum* 121.1 (2004): 93-100.

Sauer, Thomas J., and Robert Horton. "Soil heat flux." (2005): 131.

Sedighi, Majid, et al. "Energy balance at the soil atmospheric interface." *Environmental Geotechnics* 5.3 (2016): 146-157.

Shin, Jinho, et al. "A study on statistical downscaling for projection of future temperature change simulated by ECHO-G/S over the Korean peninsula." *Atmosphere* 19.2 (2009): 107-125.

Siami-Namini, Sima, Neda Tavakoli, and Akbar Siami Namin. "A comparison of ARIMA and LSTM in forecasting time series." *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018.

Statistics KOREA, Farm households and farm households, 2021 [in. Korean]

Statistics KOREA, The results of the 2020 arable land survey, 2021 [in. Korean]

Sungkwan Youm, SungKwang Hong, and Wan-Ki Koh. "The Smart Outdoor Cultivation System using Internet of Things" *Journal of the Korea Convergence Society* (2018): 63-68.

Tenge, A. Jaihura, et al. "Diurnal soil temperature fluctuations for different erosion classes of an oxisol at Mlingano, Tanzania." *Soil and Tillage Research* 49.3 (1998): 211-217.

Yun, Hong-Bae, et al. "Effect of different colored polyethylene mulch on the change of soil temperature and yield of chinese cabbage in autumn season." *Korean Journal of Soil Science and Fertilizer* 45.4 (2012): 511-514.

Zhou, Xinxing, et al. "Air pollutant concentration prediction based on GRU method." *Journal of Physics: Conference Series*. Vol. 1168. No. 3. IOP Publishing, 2019.

## Abstract

# Development and accuracy evaluation of field soil temperature prediction model by depth using artificial intelligence and meteorological parameters

Daehyun Kim  
Major in Biosystems Engineering  
Department of Biosystem Engineering  
The Graduate School  
Seoul National University

Precise agricultural technology that analyzes weather, soil environment and crop information using various advanced technologies, minimizes resource use, and maximizes productivity, needs to be applied in the field, but it is difficult to control the environment due to the influence of external weather. In the case of this, the efficiency is low compared to the cost of the equipment, so there are few cases where it is applied to the actual agricultural environment.

Of these, soil is directly involved in crop growth, so it is essential to acquire soil environment data for the normal growth of crops. Because it is important information, it is necessary to acquire data

according to depth.

In this study, as a way to acquire soil temperature data by depth without the cost of measuring equipment, we made a model that predicts soil temperature by depth over time using artificial intelligence, and compares the predicted value derived through the model with the actual value. and, it was reviewed whether it can be applied to actual agriculture.

Time unit geothermal data at three depths of 10 cm, 20 cm, and 30 cm using three machine learning models, RNN, LSTM, and GRU, and five meteorological factors: temperature, humidity, wind direction, wind speed, and insolation, which were judged to have an influence on the geothermal temperature. The optimal combination of parameters and models to accurately predict Confirmed. And, the applicability to actual agricultural activities was judged by confirming the similarity of the predicted values made from the two combinations evaluated as having the best performance through pattern analysis and residual analysis with the actual values. Evaluation was made through the values of three indicators: Root Mean Square Error (RMSE), Nash - Sutcliffe Efficiency (NSE), and Determination Coefficient ( $R^2$ ).

As a result of the experiment, the combination of parameters and models using temperature, humidity, wind speed, solar radiation, average monthly temperature data and LSTM model showed the highest performance. and temperature, humidity, wind speed, insolation, soil temperature data by depth in other regions, and the parameter and model combination using the RNN model showed the second highest performance. In addition, as a result of residual analysis with the measured values, it was confirmed that the residuals were normal and the outlier rate at all depths was around 1%.

Both the predicted values of the two datasets and model combinations selected through the experimental process were judged to



show high agreement with the actual values through evaluation index, pattern analysis, and residual analysis, and were evaluated as applicable to actual agricultural activities.

**Keywords :** Soil temperature, field, artificial intelligence, time series prediction.

*Student Number :* 2020-25115