

기후변화에 따른 목초생산량예측을 위한 회귀모델 연구*

신문선, 황선민*, 이성원**
 건국대학교 컴퓨터공학과, *비즈포스(주), **충북대학교 정보통계학과
 e-mail: msshin@kku.ac.kr

A study on regression models for predicting pasture production according to climate change

Moonsun Shin, Seonmin Hwang, Seongwon Lee
 Dept. of Computer Engineering, Konkuk University
 *Bizforce Co.,LTD,
 **Dept. of Information Statistics, Chungbuk University,

요약

한반도의 기후변화는 극한 기온으로 변화되고 있으며 이러한 이상기후로 축산농가의 사료작물 생산량에도 피해가 증가하고 있다. 사료작물의 생산량 및 재배환경 변화 등에 기후변화에 따른 영향 취약성 평가가 필요하며, 정부에서는 매년 취약성평가와 함께 조사료의 안정적인 확보와 생산성 저하 문제를 해결하기 위해 노력하고 있다. 본 연구에서는 사료작물 중 다년간의 목초 생산량 데이터와 기후데이터를 수집하여 기후변화에 따른 목초 사료작물의 생산량변화를 분석하고 예측하기 위한 연구를 수행하였다. 목초의 생산지역, 생산년도에 따른 생산량 데이터와 해당연도 지역의 기후데이터를 수집하여 사료작물 생산량 예측모델을 위한 기초데이터를 구축하였다. 또한 다양한 회귀분석을 적용하여 생산량을 예측할 수 있는 예측모델을 분석 하여 생산량에 영향을 미치는 주요 기후 요소들을 판별하였다.

1. 서론

전세계적으로 이상기후가 발생하고 있으며 기후변화에 따른 축산분야의 취약성 평가를 위한 지표의 개발 및 적용에 관한 연구와 대응정책이 정부 차원에서 이루어지고 있으나 여전히 사료작물에 대한 취약성 평가 연구와 대응정책 수립은 미흡하다. 따라서 기후변화에 따른 축산·사료작물의 적지 및 생산성 변화에 관한 영향·취약성 평가지표 선정과 방법론의 구축이 필요하다. 본 연구에서는 목초 사료작물의 기후 영향 취약성 분석을 위한 생산량 예측모델 분석을 통해 생산량에 영향을 미치는 기후요소를 판별한다.

2. 목초생산량데이터 기초 DB 구축

기초 데이터셋은 목초 생산성 실태조사를 통해 확보한 목초 생산량(건물 수량, kg/ha)과 기상자료 개방 포털의 종관기상 관측(ASOS) 데이터를 수집하였으며, 이를 건물 수량에 따라 연도별, 지역별로 구분하여 N=601인 기초데이터 DB를 구축하였다. 기상자료 개방 포털의 종관기상관측(ASOS)에서 지역별 연 단위로 GDDTOTAL(1월부터 12월까지 유효적산온

도), 월별 평균기온, 강수량, 최고기온, 최저기온 등을 기후지표로 선정하여 기초데이터를 수집하고 건물수량 데이터와 함께 분석할 기초데이터 DB를 구축하였다. [그림 1]은 구축된 기초데이터 DB이다.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	DMYTOTAL	GDDTOTAL	DDAYS	MAXTJUL	MAXTAUG	MAXTMAR	MAXTNOV	MTJAN	MTFEB	MITMAR	PREDJUL	PREDNOV	MINTJAN	MINTFEB	MINTMAR
2	13330	3201.3	1	31.1	30.3	9.8	9.6	-0.5	-0.2	3.6	13	14	-4.7	-5	-2.4
3	12120	3201.3	1	31.1	30.3	9.8	9.6	-0.5	-0.2	3.6	13	14	-4.7	-5	-2.4
4	12660	3201.3	1	31.1	30.3	9.8	9.6	-0.5	-0.2	3.6	13	14	-4.7	-5	-2.4
5	9930	3201.3	1	31.1	30.3	9.8	9.6	-0.5	-0.2	3.6	13	14	-4.7	-5	-2.4
6	13180	3201.3	1	31.1	30.3	9.8	9.6	-0.5	-0.2	3.6	13	14	-4.7	-5	-2.4
7	9750	3038.7	4	27.5	29.7	8.4	10.7	-4.6	-2.9	2.9	12	6	-10.2	-7.7	-2.1
8	9340	3038.7	4	27.5	29.7	8.4	10.7	-4.6	-2.9	2.9	12	6	-10.2	-7.7	-2.1
9	6310	3038.7	4	27.5	29.7	8.4	10.7	-4.6	-2.9	2.9	12	6	-10.2	-7.7	-2.1
10	6340	3038.7	4	27.5	29.7	8.4	10.7	-4.6	-2.9	2.9	12	6	-10.2	-7.7	-2.1
11	6790	3038.7	4	27.5	29.7	8.4	10.7	-4.6	-2.9	2.9	12	6	-10.2	-7.7	-2.1
12	6630	3038.7	4	27.5	29.7	8.4	10.7	-4.6	-2.9	2.9	12	6	-10.2	-7.7	-2.1
13	4840	3038.7	4	27.5	29.7	8.4	10.7	-4.6	-2.9	2.9	12	6	-10.2	-7.7	-2.1
14	4840	3038.7	4	27.5	29.7	8.4	10.7	-4.6	-2.9	2.9	12	6	-10.2	-7.7	-2.1
15	3890	3038.7	4	27.5	29.7	8.4	10.7	-4.6	-2.9	2.9	12	6	-10.2	-7.7	-2.1
16	8188	3577.4	5	29.3	30.4	13.3	13.5	-1.1	1	5.8	12	15	-5.8	-3.9	-1.5
17	6207	3577.4	5	29.3	30.4	13.3	13.5	-1.1	1	5.8	12	15	-5.8	-3.9	-1.5
18	6650	3577.4	5	29.3	30.4	13.3	13.5	-1.1	1	5.8	12	15	-5.8	-3.9	-1.5
19	5527	3577.4	5	29.3	30.4	13.3	13.5	-1.1	1	5.8	12	15	-5.8	-3.9	-1.5

[그림 1] 목초생산량 기초데이터셋DB

목초의 생산량에 영향을 미치는 기후요소로는 유효적산온도, 10일간 비가 오지않는 경우를 카운트한 가뭄일수 DDDAYS, 여름과 겨울의 최저기온, 최고기온, 8월과 11월 강수량 등이며 [표 1]에서 선정된 기후요소를 보여준다.

* 본 연구는 농촌진흥청 신농업기후변화대응체계구축(R&D)사업(과제번호 RS-2024-00400758)의 지원으로 수행되었음.

[표 1] 선정된기후요소

분류	지표	설명
강수	PREDAUG	8월강수량
	PREDNOV	11월강수량
기온	GDDFJTA	1월-12월 유효적상온도 (일최고기온-일최저기온-평균) - 5
	DDAYS	가뭄일수
	MAXTMAR	3월최고기온
	MAXTNOV	11월최고기온
	MAXTJUL	7월최고기온
	MAXTAUG	8월최고기온
	MINTJAN	1월최저기온
	MINTFEB	2월최저기온
	MINTMAR	3월최저기온
	MTJAN	1월평균기온
	MTFEB	2월평균기온
MTMAR	3월평균기온	
생산량	DMYT	건물수량

건물수량을 종속변수로, 선정된 기후데이터를 독립변수로 기술통계량은 [그림 2]와 같으며 정규성을 만족하는 것으로 확인되었다.

종속변수 DMYTOTAL 기술통계

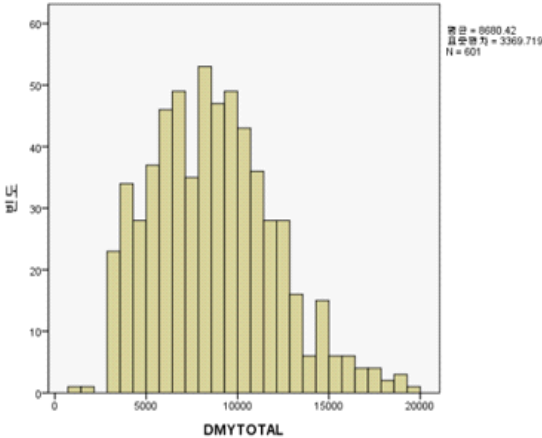
		통계량	표준오차
DMYTOTAL	평균	8680.42	137.454
	평균의 95% 신뢰구간	8410.47	
	하한	8950.37	
	상한		
	5% 절사평균	8546.79	
	중위수	8437.00	
	분산	11355009.018	
	표준편차	3369.719	
	최소값	957	
	최대값	19675	
	범위	18719	
	사분위수 범위	4709	
	왜도	.496	.100
	첨도	.009	.199

정규성 검정

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	통계량	자유도	유의확률	통계량	자유도	유의확률
DMYTOTAL	.047	601	.008	.979	601	.000

a. Lilliefors 유의확률 수정

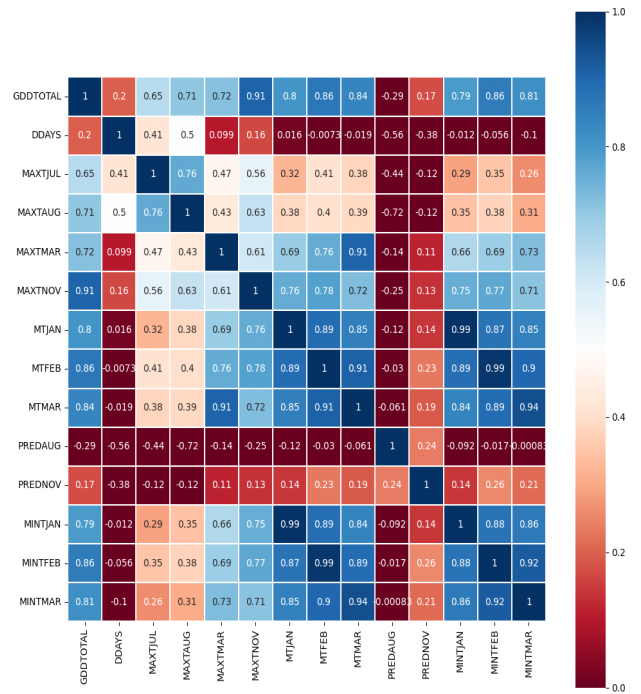
히스토그램



[그림 2] 종속변수(DMYTOTAL)기술통계량 및 정규성검증

3. 목초 생산량예측을 위한 회귀분석

목초 생산성과 기후데이터들간의 영향을 분석하기 위해서 종건물수량을 종속변수로 기후요소들을 독립변수로 단계선택법에 의한 다변량 회귀분석을 수행하기 전에 독립변수와 종속변수간의 상관분석을 수행하였다. [그림 3]은 상관분석 결과를 시각화한 히트맵이다. 기후요소들간의 연관성이 높아서 7월최고기온과 8월최고기온 혹은 1월최저기온과 2월최저기온 등과 같은 기후요소들은 모두 포함시키지 않아도 의미가 있는 것으로 해석된다.



[그림 3] 상관분석결과 시각화

다변량 회귀 분석에서 단계선택법을 적용하여 독립변수들중 종건물수량에 영향을 미치는 주요 변수들을 도출하였다. [그림 4] 최종모형에서는 7개의 변수가 포함되었으며 R²의 값은 0.274, 수정된 R²의 값은 0.266 인 것으로 확인되었다.

모형 요약

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차
1	.379 ^a	.144	.142	3120.519
2	.428 ^b	.183	.180	3050.983
3	.478 ^c	.228	.224	2967.534
4	.494 ^d	.244	.239	2939.434
5	.505 ^e	.255	.249	2919.796
6	.518 ^f	.268	.261	2897.610
7	.524 ^g	.274	.266	2887.219

- a. 예측자: (상수), MTMAR
- b. 예측자: (상수), MTMAR, MAXTAUG
- c. 예측자: (상수), MTMAR, MAXTAUG, MAXTJUL
- d. 예측자: (상수), MTMAR, MAXTAUG, MAXTJUL, DDAYS
- e. 예측자: (상수), MTMAR, MAXTAUG, MAXTJUL, DDAYS, INTMAR
- f. 예측자: (상수), MTMAR, MAXTAUG, MAXTJUL, DDAYS, INTMAR, MTJAN
- g. 예측자: (상수), MTMAR, MAXTAUG, MAXTJUL, DDAYS, MINTJAN, MINTFEB, PREDAUG

[그림 4] 모형요약

설명계수인 R²의 값이 다소 낮은 편이며 이는 목초사료작물

의 특성상 초지 조성후 1년내내 기후의 영향을 받는 사료작물로 이탈리아라이그라스와 같은 동계사료작물과는 다르게 기후요소들과의 연관성이 밀집되지 않음을 의미하는 것으로 분석된다. 그럼에도 불구하고 7-8월의 최고기온과 8월 강수량에 영향을 비교적 받고 있음을 알 수 있다.

그리고 10일이상 연속적으로 비가 오지 않는 날들을 계산한 가뭄일수 기후지표인 DDAY5도 최종모형에 포함되어 있어 영향을 받고 있음을 시사한다. 그러나 ANOVA테이블과 [그림 5]의 단계별 계수를 확인한 결과 3월평균기온(MTMAR) 변수와 3월최저기온(MINTMAR) 변수는 공선성계수통계량인 VIF 값이 10이상으로 다중공선성이 있는 것으로 확인되어 두 변수 중 한 변수를 제외하고 예측모형을 적용할 수 있다.

순번	변수	계수*				유의확률	공선성 통계량	
		B	표준오차	t	표준화 계수		공차	VIF
1	(상수)	6200.290	278.050	22.299	.000			
	MITMAR	497.251	49.563	.379	10.033	1.000	1.000	
2	(상수)	14614.444	1596.264	9.155	.000			
	MITMAR	605.133	52.559	.462	11.532	.850	1.176	
	MAXTAUG	-308.486	57.668	-.214	-5.349	.000	.850	
3	(상수)	8830.357	1834.007	4.815	.000			
	MITMAR	562.899	51.640	.429	10.900	.833	1.200	
	MAXTAUG	-655.243	81.064	-.456	-8.083	.000	.407	
	MAXTJUL	564.882	95.339	.333	5.925	.000	.410	
4	(상수)	6235.241	1959.672	3.182	.002			
	DDAYS	-272.386	54.357	-.182	-4.232	.000	.663	
	MINTMAR	-496.686	131.475	-.446	-3.778	.000	.089	
	MITJAN	189.257	59.416	.224	3.185	.002	.249	
7	(상수)	12001.188	2950.898	4.053	.000			
	MITMAR	845.863	157.064	.646	5.392	.000	.085	
	MAXTAUG	-704.678	110.679	-.490	-6.367	.000	.207	
	MAXTJUL	590.271	99.327	.348	5.943	.000	.357	
	DDAYS	-314.341	56.683	-.211	-4.714	.000	.614	
	MINTMAR	-475.037	131.342	-.426	-3.617	.000	.088	
	MITJAN	179.190	59.364	.212	3.018	.003	.247	
	PREDAUG	-110.674	48.180	-.131	-2.299	.022	.375	

[그림 5] 모형 계수 및 VIF 확인

다중회귀에서 발생한 다중 공선성의 문제를 해결하기 위해 라소모형 적용하여 결과를 확인하였다. 라소 회귀는 다중공선성 문제를 해결하고 모형을 단순화하는 장점이 있으며 변수 선택 기능을 통해 유의미한 변수들을 식별하고, 불필요한 변수들을 제거하여 모형의 해석력과 일반화 성능을 향상시키는 선형회귀의 한 종류이다.

다변량 회귀 분석의 경우 다중 공선성의 문제가 발생하여 이를 해결하기 위해 LASSO 모델을 적용하였으며 파이썬을 활용하여 예측모형을 도출하였다. [그림 6]은 파이썬에서 LASSO 모델을 적용하여 목초 생산량 예측모형을 구하는 결과를 보여준다. 또한 설명력 계수인 수정된 R²의 값은 0.35정도로 확인되었으며 조금 더 향상된 설명력 계수 값을 확인할 수 있었다.

[그림 7]은 LASSO 모델의 테스트데이터셋에 대한 예측결과를 나타낸 그래프이다.

```

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import tensorflow as tf

from sklearn.linear_model import LinearRegression, Ridge, Lasso
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

from scipy import stats
from statsmodels.formula.api import ols, glm
herb_all = pd.read_csv('/content/drive/MyDrive/herb2024.csv', sep=',')

X = herb_all[['GDDTOTAL', 'DDAYS', 'MAXTJUL', 'MAXTAUG', 'MAXTMAR', 'MAXTNOV',
             'MTJAN', 'MTFEB', 'MTMAR', 'PREDAUG', 'PREDNV', 'MINTJAN', 'MINTFEB', 'MINTMAR']]

Y = herb_all['DMYTOTAL']
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, random_state=1000)
    
```

```

# Lasso Regression
lasso_reg = Lasso(alpha=0.3, max_iter=10000)
lasso_reg.fit(X_train, Y_train)
lasso_pred = lasso_reg.predict(X_test)
lasso_r2 = r2_score(Y_test, lasso_pred)
print('Lasso R2 Score : ', lasso_r2)
    
```

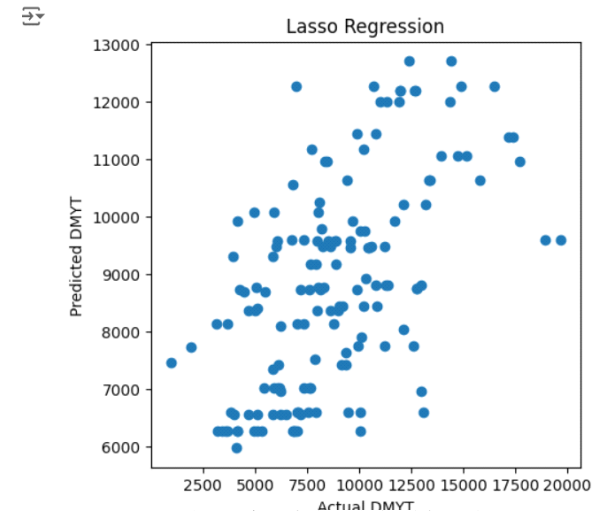
Lasso R2 Score : 0.3350374450213778

[그림 6] 라소회귀 실험 결과

```

plt.scatter(Y_test, lasso_pred)
plt.xlabel("Actual DMYT")
plt.ylabel("Predicted DMYT")
plt.title("Lasso Regression")

plt.show()
    
```



[그림 7] 라소회귀모형의 예측결과

[그림 7]의 그래프에서 보여지듯이 목초생산량예측모형의 설명계수가 0.35로 다소 낮은 편이어서 실제값과 예측값의 결과 편차가 다소 큰 것으로 보여진다.

4. 결론

본 연구에서는 기후변화에 따른 영향취약성을 분석하기 위해 목초 사료작물의 생산량 데이터와 연관 기후데이터를 수집하여 다양한 회귀모형을 적용한 목초생산량예측모형을 분석하였다. 목초 생산량 데이터와 해당연도 지역의 기후데이터를 수집하였으며 분석을 위한 기초데이터 셋을 가공하여 단계선택법 다중회귀분석과 라소 회귀모형을 적용한 목초 생산량

예측모형을 구축하여 생산량에 영향을 미치는 주요 기후 요소들을 분석하였으며 분석 결과 목초의 경우 강우량의 영향력이 큰 것으로 확인되었다. 도출된 목초 생산량예측모형의 경우 독립변수가 다소 많은 것으로 판단되며 독립변수의 개수가 적은 모형이 기후요소의 영향 취약성 설명에 적합할 것으로 분석되었다. 따라서 향후 다양한 독립변수 구성으로 예측모형을 구하고 결과를 분석하는 실험 수행이 추가적으로 요구된다.

Acknowledgement

본 연구는 농촌진흥청 신농업기후변화대응체계구축(R&D)사업(과제번호 RS-2024-00400758)의 지원으로 수행되었음.

참고문헌

- [1] Hatfield, J.L., Morton, L.W. and Hall, B. 2018. Vulnerability of grain crops and croplands in the Midwest to climatic variability and adaptation strategies. *Climatic Change* January. 146:263-275.
- [2] Jeong Sung Jung, Hyung Soo Park, Hee Jung Ji, Ki Yoong Kim, Se Young Lee and Bae Hun Lee, Predicting Changes in the Suitable Agro-climate Zone of Italian Ryegrass Cultivars with RCP 8.5 Climate Change Scenario, *Journal of the Korean Society of Grassland and Forage Science* 2020 40(4): pp.196~204, <https://doi.org/10.5333/KGFS.2020.40.4.196>
- [3] Impact of Climate Change on Agriculture and Its Mitigation Strategies: A Review. *Sustainability* 2021, 13, 1318. <https://doi.org/10.3390/su13031318>
- [4] Stöckle, C.O., Higgins, S., Nelson, R., Abatzoglou, J., Huggins, D., Pan, W., Karimi, T., Antle, J., Eigenbrode, S.D. and Brooks, E. 2018. Evaluating opportunities for an increased role of winter crops as adaptation to climate change in dryland cropping systems of the US Inland Pacific Northwest. *Climatic change* 146: 247-261.
- [5] M.S.Shin, S.W.Lee, S.M.Hwang and S.H.Moon, "Correlation Analysis of Production Changes of Feed Crop Data according to Climate Change" , in *Proceeding of the 14th Annual International Conference on Future Information and Communication Engineering* , Jeju: pp. 149-152, 2022.