

Development of a soil total carbon prediction model using a multiple regression analysis method

Jun-Hyuk Yoo^{1,†}, Jwa-Kyoung Sung^{2,†}, Deogratus Luyima¹, Taek-Keun Oh^{1,*}, Jaesung Cho^{3,*}

¹Department of Bio-environmental Chemistry, Chungnam National University, Daejeon 34134, Korea

²Department of Crop Science, Chungbuk National University, Cheongju 28644, Korea

³Department of Animal Science and Biotechnology, Chungnam National University, Daejeon 34134, Korea

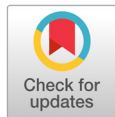
[†]This author contributed to this study as first author.

^{*}Corresponding authors: ok5382@cnu.ac.kr, kor0025cho@cnu.ac.kr

Abstract

There is a need for a technology that can quickly and accurately analyze soil carbon contents. Existing soil carbon analysis methods are cumbersome in terms of professional manpower requirements, time, and cost. It is against this background that the present study leverages the soil physical properties of color and water content levels to develop a model capable of predicting the carbon content of soil sample. To predict the total carbon content of soil, the RGB values, water content of the soil, and lux levels were analyzed and used as statistical data. However, when R, G, and B with high correlations were all included in a multiple regression analysis as independent variables, a high level of multicollinearity was noted and G was thus excluded from the model. The estimates showed that the estimation coefficients for all independent variables were statistically significant at a significance level of 1%. The elastic values of R and B for the soil carbon content, which are of major interest in this study, were -2.90 and 1.47, respectively, showing that a 1% increase in the R value was correlated with a 2.90% decrease in the carbon content, whereas a 1% increase in the B value tallied with a 1.47% increase in the carbon content. Coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE) methods were used for regression verification, and calibration samples showed higher accuracy than the validation samples in terms of R^2 and MAPE.

Keywords: agricultural land, multiple regression analysis, prediction model, RGB values, soil total carbon



OPEN ACCESS

Citation: Yoo JH, Sung JK, Luyima D, Oh TK, Cho J. Development of a soil total carbon prediction model using a multiple regression analysis method. Korean Journal of Agricultural Science 48:891-897. <https://doi.org/10.7744/kjoas.20210075>

Received: September 14, 2021

Revised: October 27, 2021

Accepted: November 11, 2021

Copyright: © 2021 Korean Journal of Agricultural Science



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

토양은 식생에 영양분을 공급하고 생명체의 순환을 돕는 필수적인 환경요소이며, 토양 유기물은 토양의 질을 향상시키는 중요한 역할을 수행한다(McBratney et al., 2014; Jeong, 2016). 토양 유기물의 주요한 구성성분인 토양 탄소는 토양의 용적밀도, 수분함량, 양이온교환용량 등과 밀접한 관련성을 가지고 있어 토양의 질에서 가장 중요한 요소로 언급되고 있다(Schoenholtz et

al., 2000; Lee et al., 2018; Yoo et al., 2021). 특히, 토양의 탄소함량은 토양의 비옥도에 직접적인 영향을 미쳐 작물의 생육과 수확량에 큰 영향을 미치는 것으로 알려져 있으며, Zvomuya 등(2008)의 연구에 따르면 토양 유기 탄소(soil organic carbon, SOC) 함량이 증가함에 따라 작물의 생산량이 한계점까지 비례적으로 증가하는 경향이 나타났다고 보고하였다. 또한 현대사회에서 온실가스 배출량 증가에 따른 기후변화에 대한 불안감이 증가하고 있는 가운데, 토양에 탄소를 장기간 격리함으로써 온실가스 배출을 완화할 수 있는 방안이 연구되고 큰 관심을 받고 있다(Bell and Worrall, 2009; McBratney et al., 2014; Choi et al., 2017; Oh et al., 2017). 자연의 탄소 순환 과정에서 토양은 지구상의 유기 탄소를 가장 많이 저장하기 때문에 거대한 탄소 저장고 역할을 수행하고 있으며(Lal, 2008; Han et al., 2016), 전 세계에 존재하는 탄소의 약 60 - 70%를 포함하고 있다(Brady and Weil, 2008; Smith, 2012).

위와 같이 토양 탄소는 토질, 작물생산량 및 기후변화 완화에 중요한 역할을 수행하기 때문에 토양 탄소함량을 신속하고 정확하게 분석하는 기술이 요구되어 왔다. 현재 토양 탄소함량을 분석하는 대표적인 방법으로는 Tyurin 법, Walkley-Black 법과 같이 산화제인 중크롬산을 사용하여 분석하는 습식산화법과 분석 기기 내에 있는 고온의 회화로에서 시료를 연소시켜 분석하는 방법인 건식연소법 등이 있다(Cho et al., 2013). 습식산화법은 강력한 산화제인 중크롬산을 이용하여 유기탄소를 산화시켜 정량하는 방법으로 유독성 화합물인 중크롬산칼륨과 진한 황산을 사용하기 때문에 시험조작 시 위험할 수 있고 중금속을 함유하고 있는 다량의 폐액이 발생하여 처리에 각별한 주의가 기울여야 한다(Seo et al., 2004). 건식연소법은 고가의 장비가 필요하고 장비를 다룰 수 있는 전문 인력이 필요하다는 단점이 있다. 이에 따라 전문적인 인력, 시간, 비용이 많이 소요되고 위해성 및 환경오염 유발 가능성을 가지고 있는 기존의 토양 탄소함량 분석법의 단점을 최소화한 새로운 분석 방법 개발이 필요하다고 판단된다.

토양의 물리적 특성 중 토양색은 토양의 탄소함량과 밀접한 관련을 갖는데, 토양색에 영향을 끼치는 주요한 요인으로는 조도(Lx), 수분함량(water content), 유기물함량 등이 있다(Chung et al., 2006). 토양 유기물은 암갈색 내지 흑색을 띠고 있어 토양에 유기물이 많은 토양일수록 어두운 색을 띠는 특징을 가지고 있다(Chung et al., 2006). 토양의 색깔은 수분함량에 따라서도 크게 달라지는데, 수분이 많은 토양은 어두운 색을 나타내고 수분이 증발하면 밝은 색으로 변하는 특징을 가지고 있다. 또한 토양의 유기물 함량을 통해 토양 수분함량을 예측하는 연구가 이루어지고 있는 만큼 토양의 수분함량과 탄소함량은 깊은 연관성을 가지고 있다(Hur et al., 2014).

따라서 본 연구는 토양의 총 탄소함량(total carbon, T-C)을 예측하는 회귀모델을 개발하기 위하여 충청남도 농경지에서 채취한 토양의 실내 사진 촬영을 통해 도출한 R (red), G (green), B (blue) 값과 토양의 수분함량, 실내 조도를 측정된 값을 토대로 데이터를 정리하고 통계분석 프로그램을 통한 다중회귀분석을 실시하였고, 예측 모델의 정확성을 평가하기 위해 산출된 회귀모델을 검증하였다.

Materials and Methods

토양 채취 지역 및 데이터 취득 방법

토양 샘플은 2020년 8월부터 2020년 9월까지 충청남도(35°58'N-37°03'N 125°32'E-127°38'E) 4개 시군(공주시, 금산군, 논산시, 부여군)의 18지점에서 농경지의 종류에 따라 논 토양 8지점, 밭 토양 7지점, 과수원 토양 3지점을 채취하였다(Fig. 1). 토양 불교란 시료채취기(gouge auger)를 사용하여 농경지 토양을 채취하였고, gouge auger에 채취된 토양의 수분 증발 방지를 위해 랩으로 밀봉하여 실험실로 운반하고 분석을 실시하였다.

자체 제작한 격자틀(격자 간격 2 cm)을 gouge auger 위에 올려놓음으로써 토양 단면의 구역을 구분하였고, 사진 촬영 전 격자틀 높이에서 실내 조도를 측정하였다. 그리고 토양 단면을 실제와 같은 색상으로 촬영하고 기술의 호환성을 확보하기 위해 스마트폰(SM-G973N [S10], Samsung electronics, Suwon, Korea)의 기본 카메라 어플리케이션

(원본 필터)으로 촬영하였다. 촬영된 사진은 색상 분석 프로그램(Color Cop, v5.4.6, (c) Jay Prall, Massachusetts, USA)을 사용하여 토양 단면의 RGB 값을 분석하였다. 채취한 토양의 수분함량은 forced convection oven을 105°C, 24시간 설정하여 건조 전과 후의 토양 중량 차이를 통해 계산하였고, 토양의 총 탄소함량은 건조된 토양을 2 mm sieve로 채질하여 CHN analyzers (TrueSpec CHN, Leco Co., Michigan, USA)를 통해 분석하였다.

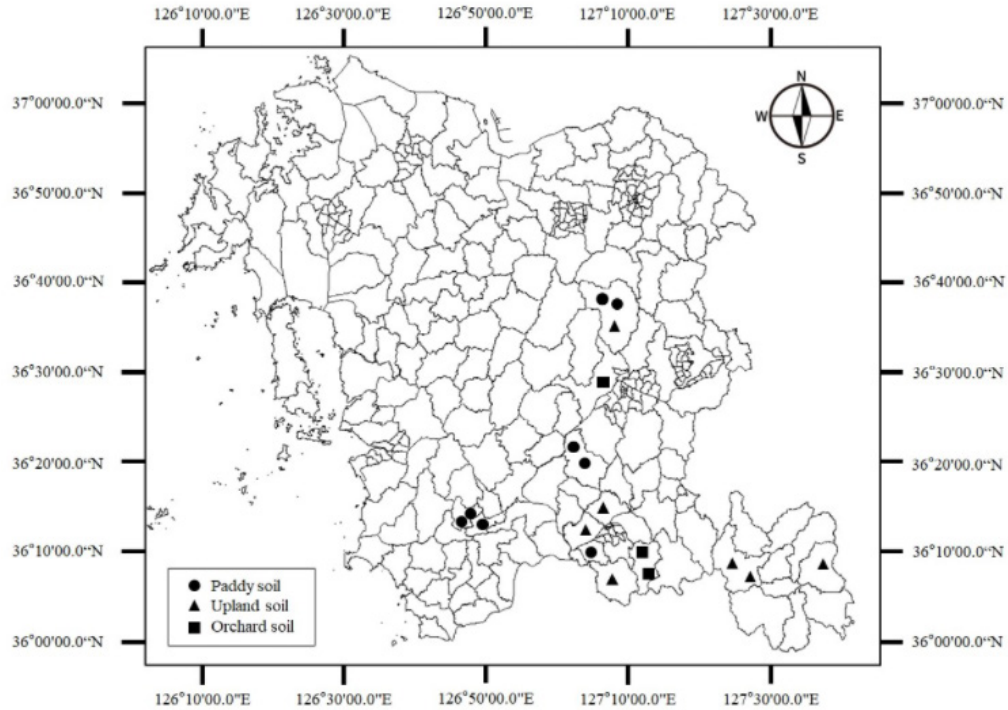


Fig. 1. Location of soil sampling point in Chungcheongnam-do province, Republic of Korea.

통계 분석 방법

토양의 총 탄소함량을 예측하기 위해 토양 단면의 RGB 값과 수분함량, 조도를 분석 데이터로 사용하였으며, 통계 분석에 사용된 변수들의 기술 통계량은 Table 1과 같다. 종속변수는 토양의 총 탄소함량이며, 나머지 변수는 모두 독립변수이다. 독립변수 중 RGB 값을 제외한 조도와 수분함량은 통제변수로 모형에 포함하였다. 조도를 제외한 모든 변수는 연속형 변수이며, 분석 자료에서 조도가 갖는 값은 8개에 불과하여 범주형 변수로 분석에 사용하였다.

Table 1. Basic statistics used in multiple regression analysis. (n = 378)

Classification	T-C (%)	R (red)	G (green)	B (blue)	Water content (%)	lx (lm ⁻²)
Average	0.90	154.15	130.87	110.00	22.05	277.11
Standard deviation	0.80	37.88	28.29	24.37	6.77	66.32
Minimum value	0.04	47.00	37.00	31.00	9.29	182.00
Maximum value	3.01	241.00	198.00	179.00	45.13	376.00

T-C, total carbon.

본 연구의 목적은 RGB 값과 토양의 총 탄소함량의 관계를 살펴보는 것이기 때문에 우선 주요 관심 변수인 R, G, B, 토양 총 탄소함량의 관계를 상관계 분석을 통해 살펴본 후 로그-로그 형태의 선형 회귀모델을 최소제곱법 (least-squares estimation)과 강건표준오차(robust standard error)를 이용하여 추정함으로써 RGB 값이 토양의 총 탄소함량에 미치는 영향을 탄성치(elasticity) 형태로 분석하였다. 통계 분석에 사용한 프로그램은 Stata (StataCorp. version 16, Texas, USA)이다.

회귀식 검증 방법

회귀식 도출을 위해 378개(62%)의 학습 데이터(training data)를 사용하였으며, 회귀식 검증을 위해 231개(38%)의 검증 데이터(validation data)를 사용하였다. 또한 산출된 회귀식을 다양한 측면에서 검증하기 위해 결정계수 (coefficient of determination, R^2), 평균 제곱근 오차(root mean square error, RMSE), 평균 절대비 오차(mean absolute percentage error, MAPE) 방법을 사용하여 training data와 validation data에서의 예측력을 검정하였다.

Results and Discussion

상관관계 분석 결과

상관관계 분석 결과, R과 G와 B는 모두 유의수준 5% 내에서 통계적으로 유의하게 토양의 총 탄소함량과 부(-)의 관계를 갖는 것으로 나타났으며, RGB 간에는 매우 높은 수준의 양(+)의 상관관계가 존재하는 것으로 나타났다 (Table 2). 특히, R과 G, G와 B의 상관계수는 각각 0.90과 0.92로 나타나 RGB 값을 모두 독립변수로 포함한 선형 회귀모델을 추정할 경우 높은 수준의 다중공선성 문제가 발생할 것으로 추정되었다.

Table 2. Correlation analysis results of T-C, R, G, and B.

Variable	T-C	R	G	B
T-C	1.00			
R	-0.40**	1.00		
G	-0.25**	0.90**	1.00	
B	-0.11*	0.71**	0.92**	1.00

T-C, total carbon; R, red; G, green; B, blue.

** , * indicate statistically significant at 1, 5% significance levels, respectively.

선형회귀 분석 결과

상관관계 분석은 단순히 두 변수 간의 관계를 분석하는 것으로 다수의 변수가 종합적으로 특정 변수에 미치는 영향을 파악할 수 없다. 따라서 본 연구에서는 상관관계 분석에 이어 로그-로그 선형 회귀모델을 이용한 다중회귀 분석을 수행하였다. 로그-로그 선형 회귀모델의 종속변수는 토양의 총 탄소함량이며, 독립변수는 R, G, B, 수분함량, 조도이다. 앞서 분석자료에서 언급한 것처럼 조도를 제외한 모든 변수는 연속형 변수로 로그 치환하여 분석에 사용하였고, 조도는 범주형 변수임으로 로그 치환없이 분석에 사용하였다.

회귀모델의 추정 결과, 상관관계 분석에서 예상하였듯이 R, G, B를 모두 독립변수로 모델에 포함한 경우, 높은 수준의 다중공선성 문제가 발생하였다. 따라서 나머지 두 변수와 높은 관계성을 가진 G를 모델에서 제외한 후 모델을 재추정하였으며, 그 결과는 Table 3와 같다. 그 결과, 모든 독립변수들의 추정 계수(탄성치)는 유의수준 1%에서 통계적으로 유의하였다. 토양 탄소함량에 대한 R과 B의 탄성치는 각각 -2.90과 1.47로 나타나 R 값이 1% 증가하면 탄

소함량은 2.90% 감소하고, B 값이 1% 증가하면 탄소함량은 1.47% 증가하는 것으로 나타났다. 또한 수분함량도 R과 B에 비하면 작지만 탄소함량에 대한 수분함량의 탄성치는 0.58로 수분함량이 1% 증가하면 탄소함량은 0.58% 증가하는 것으로 나타났다. 또한 조도가 갖는 8개의 값에서 산출된 추정 계수를 이용하여 회귀식에 각각 대입할 경우, 예측 모델의 범용성 저하 문제가 발생할 것으로 판단되어 조도가 갖는 8개(182 - 376 $\text{lm}\cdot\text{m}^{-2}$) 추정 계수의 평균(1.37)을 회귀식에 대입하였다. 다중회귀분석을 통해 산출한 토양의 총 탄소함량 예측 회귀식은 식(1)과 같다.

$$\ln(\text{T-C}) = 5.48 - 2.90 \ln(\text{R}) + 1.47 \ln(\text{B}) + 0.58 \ln(\text{WC}) \quad (1)$$

Abbreviations: T-C, total carbon; WC, water content.

Table 3. Multiple regression analysis results for deriving of regression models.

Variable	Estimation coefficient	Standard deviation	t-value
$\ln(\text{R})$	-2.90**	0.27	-10.89
$\ln(\text{B})$	1.47**	0.27	5.46
$\ln(\text{water content})$	0.58**	0.21	2.82
Constant	4.11**	1.19	3.44

R, red; G, green.

** , * indicate statistically significant at 1, 5% significance levels, respectively.

회귀식 검증 결과

검증 방법 중 R^2 값은 validation data이 0.76으로 training data (0.53)보다 높은 값을 나타냈지만, RMSE 값은 validation data가 0.29으로 training data (0.20)보다 높은 값을 나타내었다(Table 4). MAPE 값은 training data가 43.59로 validation data (23.87)보다 높은 값을 나타내었다. 즉, 다른 농경지의 토양을 채취하여 본 연구와 같이 RGB 값과 수분함량을 정확하게 측정하여 예측 회귀식에 대입할 경우, 위에서 검증한 RMSE의 값에 따라 오차범위 약 0.29 안에서 농경지의 총 탄소함량을 예측할 수 있을 것으로 판단된다.

예측 모델의 검증은 위에서 산출한 4개의 회귀식을 각각 검증하여 평가하였다(Table 4). 먼저 농경지 토양 전체의 데이터로 산출한 회귀식의 검증은 R^2 값이 validation data가 0.453으로 training data (0.391)보다 높은 값을 나타냈지만, RMSE 값은 validation data가 0.652으로 training data (0.607)보다 높은 값을 나타내었다. MAPE 값은 training data가 57.593으로 validation data (53.992)보다 높은 값을 나타내었다. 즉, 다른 농경지의 토양을 채취하고 본 연구와 같이 RGB 값과 수분함량을 정확하게 측정하여 예측 회귀식에 대입할 경우, 위에서 검증한 RMSE의 값에 따라 오차범위 약 0.652 안에서 농경지의 총 탄소함량을 예측할 수 있을 것으로 판단된다.

Table 4. Results of T-C content prediction model's validation.

Source	R^2_{tra}	RMSE _{tra}	MAPE _{tra}	R^2_{val}	RMSE _{val}	MAPE _{val}
Value	0.53	0.20	43.59	0.76	0.29	23.87

T-C, total carbon; R^2 , coefficient of determination; RMSE, root mean square error; MAPE, mean absolute percentage error; tra, training data; val, validation data.

Conclusion

본 연구는 농경지에서 채취한 토양의 사진 촬영을 통해 도출한 RGB 값과 토양의 수분함량, 조도를 측정된 데이터를 정리하고, 다중회귀분석을 실시하여 토양의 총 탄소함량을 예측하는 모델을 개발하였고 예측 모델의 정확성을 평가하였다. 상관관계 분석 결과 R, G, B 사이에서는 모두 5% 내로 토양의 총 탄소함량과 통계적으로 유의한 관

계를 나타내었고, R과 G, G와 B 사이의 매우 높은 상관관계로 인한 다중공선성 문제 때문에 회귀식 산출에서 G를 제외하였다. 토양의 총 탄소함량에 대한 R과 B의 탄성치는 각각 -2.90과 1.47로 나타났으며, 수분함량의 탄성치는 0.58로 나타났다. 회귀식 검증을 위해 R^2 , RMSE, MAPE 방법을 사용하였고, 그 결과 농경지 토양의 총 탄소함량을 예측할 경우 오차범위(RMSE) 약 0.29 안에서 결과가 도출될 것으로 예상할 수 있다. 다만 본 연구에서 처리한 토양 분석 데이터가 모든 종류의 토양을 대표할 수는 없으므로, 더욱 정확하고 신뢰성 있는 예측 모델을 개발하기 위해서는 토양의 분류를 통해 더욱 많은 데이터를 확보하여 예측 모델을 산출할 필요가 있다고 판단된다.

Conflict of Interests

No potential conflict of interest relevant to this article was reported.

Acknowledgements

본 연구는 농촌진흥청(Rural development administration)의 공동연구사업(Project No. PJ015102)의 지원으로 수행되었습니다.

Authors Information

Jun-Hyuk Yoo, <https://orcid.org/0000-0002-5861-4836>

Jwa-Kyoung Sung, <https://orcid.org/0000-0002-0758-6644>

Deogratus Luyima, <https://orcid.org/0000-0002-5133-0603>

Taek-Keun Oh, <https://orcid.org/0000-0003-0215-0427>

Jaesung Cho, <https://orcid.org/0000-0002-9720-7950>

References

- Bell MJ, Worrall F. 2009. Estimating a region's soil organic carbon baseline: The undervalued role of land-management. *Geoderma* 152:74-84. doi: 10.1016/j.geoderma.2009.05.020
- Brady NC, Weil RR. 2008. The nature and properties of soils. 14th edition. Pearson prentice hall, New jersey, USA.
- Cho MK, Kim SY, Lim YS. 2013. Method comparison of soil organic carbon measurement: Wet-oxidation & dry combustion method. *Korean Journal of Environmental Agriculture* 2013:318. [in Korean]
- Choi EJ, Lee JH, Jeong HC, Kim SH, Lim JS, Lee DK, Oh TK. 2017. Analysis of research trends in methane emissions from rice paddies in Korea. *Korean Journal of Agricultural Science* 44:463-476. doi: 10.7744/kjoas.20170055 [in Korean]
- Chung JB, Kim KH, Kim KY, Kim JG, Sa TM, Suh JS, Sohn BK, Yang JE, Eom KC, Lee SE, et al. 2006. *Soil Science*. pp. 89-93. Hyangmunsa, Seoul, Korea. [in Korean]
- Han KH, Zhang YS, Jung KH, Cho HR, Seo MJ, Sonn YK. 2016. Statistically estimated storage potential of organic carbon by its association with clay content for Korean upland subsoil. *Korean Journal of Agricultural Science* 43:353-359. doi: 10.7744/kjoas.20160037 [in Korean]
- Hur SO, Sonn YG, Hyun BK, Shin KS, Oh TK, Kim JG. 2014. Verification on PTF (Pedo-transfer function) estimating soil water retention based on soil properties. *Korean Journal of Agricultural Science* 41:391-398. doi: 10.7744/cnujas.2014.41.4.391 [in Korean]

- Jeong GY. 2016 Evaluating spectral preprocessing methods for visible and near infrared reflectance spectroscopy to predict soil carbon and nitrogen in mountainous areas. *Journal of Kgeography* 51:509-523. [in Korean]
- Lal R. 2008. Sequestration of atmospheric CO₂ in global carbon pools. *Energy & Environmental Science* 1:86-100. doi:10.1039/b809492f
- Lee JH, Seong CJ, Kang SS, Lee HC, Kim SH, Lim JS, Kim JH, Yoo JH, Park JH, Oh TK. 2018. Effect of different types of biochar on the growth of Chinese cabbage (*Brassica chinensis*). *Korean Journal of Agricultural Science* 45:197-203. doi: 10.7744/kjoas.20180033 [in Korean]
- McBratney AB, Stockmann U, Angers DA, Minasny B, Field DJ. 2014. Challenges for soil organic carbon research. In *Soil carbon* edited by Alfred E, Hartemink AE, McSweeney K. pp. 3-16. Springer, New York, USA. doi: 10.1007/978-3-319-04084-4_1
- Oh TK, Lee JH, Kim SH, Lee HC. 2017. Effect of biochar application on growth of Chinese cabbage (*Brassica chinensis*). *Korean Journal of Agricultural Science* 44:359-365. doi: 10.7744/kjoas.20170039 [in Korean]
- Schoenholtz SH, Van Miegroet H, Burger JA. 2000. A review of chemical and physical properties as indicators of forest soil quality: Challenges and opportunities. *Forest Ecology and Management* 138:335-356. doi: 10.1016/S0378-1127(00)00423-0
- Seo MC, KH So, BG Go, Sonn YK. 2004. Comparison of Tyurin method and dry combustion method for carbon analysis in soils of low inorganic carbon content. *Korean Journal of Soil Science and Fertilizer* 37:315-321 [in Korean]
- Smith P. 2012. Soils and climate change. *Current Opinion in Environmental Sustainability* 4:539-544. doi:10.1016/j.cosust.2012.06.005
- Yoo JH, Luyima D, Lee JH, Park SY, Yang JW, An JY, Yun YN, Oh TK. 2021. Effects of brewer's spent grain biochar on the growth and quality of leaf lettuce (*Lactuca sativa* L. var. *crispa*). *Applied Biological Chemistry* 64:1-10. doi: 10.1186/s13765-020-00577-z
- Zvomuya F, Janzen HH, Larney FJ, Olson BM. 2008. A long - term field bioassay of soil quality indicators in a semiarid environment. *Soil Science Society of America Journal* 72:683-692. doi: 10.2136/sssaj2007.0180