Estimation of Soil Organic Carbon (SOC) Stock in South Korea Using Digital Soil Mapping Technique

Seong-Jin Park^{1,2}, Soon-Ik Kwon¹, Seong-Heon Kim¹, Jaehong Shim¹, Yun-Hae Lee¹, and Taek-Keun Oh^{2*}

¹Researcher, Division of Soil and Fertilizer, National Institute of Agricultural Science, RDA, Wanju 55365, Korea

²Student of Dr's Degree, Department of Bio-Environmental Chemistry, College of Agriculture and Life Science, Chungnam National University, Daejeon 34134, Korea

*Corresponding author: ok5382@cnu.ac.kr

ABSTRACT

Received: March 26, 2021

Revised: March 31, 2021

Accepted: May 10, 2021

ORCID

Seong-Jin Park

<https://orcid.org/0000-0001-5102-3657>

Seong-Heon Kim

<https://orcid.org/0000-0001-6047-9103>

Taek-Keun Oh

<https://orcid.org/0000-0003-0215-0427>

Estimation of soil organic carbon (SOC) stocks is highly relevant considering that SOC is the central driver in soil fertility and climate change mitigation. This study aims to (i) evaluate the SOC stock in the first 0 - 30 cm and 0 - 100 cm soil layer on a national scale from spatially explicit explanatory environmental variables and a legacy soil database and (ii) the spatial distribution of SOC at national scale through digital mapping technique. A spatial model was established using Cubist, a decision tree algorithm and based on soil data (s factor), climatic (c factor), topographic (r factor). Results showed that soil texture, soil parent, mean annual precipitation and elevation were the most important predictors of SOC. The Cubist prediction model had a Root Mean Squared Error (RMSE) equal to 19.5 at 0 - 30 cm, 68.7 at 0 - 100 cm. The predicted mean SOC stock from fitted models was 35 ton C ha⁻¹ for 0 - 30 cm depth, 87 ton C ha⁻¹ for 0 - 100 cm soil depth. In total, soil stored approximately 330 Mt C for 0 - 30 cm depth and 842 Mt C for 0 - 100 cm depth.

Keywords: Carbon stock, Digital soil mapping, Soil organic carbon, South Korea

Uncertainty (95%) and spatial distribution map of predicted SOC stocks (ton C ha⁻¹, 0 - 30 cm soil depth) in South Korea.

Introduction

Soil organic carbon (SOC) is the essential nutrients for plant growth and major components of the global carbon cycle (Batjes, 1996). SOC affect the concentration of greenhouse gases in the atmosphere and global climate changes (Powers and Schlesinger, 2002). SOC has the ability to improve the soil physical, chemical, and biological properties, which can enhance soil security. Soil has 1,500 Gt C in the top one meter, globally (Jobbágy and Jackson, 2001) which is approximately three times as much C found in the biosphere and twice as much C found in the atmosphere. Accurate quantification of SOC is important for assessing the C sink capacity of soils and the changes rate of SOC (Yang et al., 2009).

Digital soil mapping is the creation and population of spatial soil information systems by the use of field and laboratory observational methods coupled with spatial and non-spatial soil inference systems (Lagacherie and McBratney, 2006). Digital soil mapping does not just produce a paper map; it is a dynamic process in essentially a spatial database of soil properties, based on a sample of landscape at known locations (Mcbratney et al., 2003). Field sampling is used to determine spatial distribution of soil properties, which are mostly measured in the laboratory. These data are then used to predict soil properties in areas not sampled. Spatial variability data are important for estimating the stock of soil carbon and also for understanding the biophysical processes that can influence the flow of organic carbon in soils. Digital soil maps describe the uncertainties associated with such predictions and, when based on longitudinal data can, provide information on dynamic soil properties.

Predictions of soil organic carbon stock in digital mapping are based on relationship among soil properties and environmental factor. McBratney et al. (2003) generalized and formulated a equation, with the objective of modeling the variables responsible for the processes of soil formation through an empiric quantitative description of the relationships among other environmental covariates, used as spatial prediction functions.

$$S = f(s, c, o, r, p, a, n) + \epsilon \quad (\text{Eq. 1})$$

This is the called the scorpan model and this function with following seven factors: S = soil and other properties of the soil in a given location; c = climate, climatic properties; o = organisms, vegetation, r = topography, attributes of the landscape; p = parent material, lithology; a = age, time factor; n = space, spatial location; ϵ = autocorrelated random spatial variation. Each factor is represented by a group of one or more continuous of categorical variables; for example, r for elevation, slope or other derived attribute of a DEM.

The objectives of this study were to estimate soil carbon stock at two soil depth (0 - 30 cm, 0 - 100 cm) at national scale using the digital soil mapping skill.

Materials and Methods

Study area South Korea is a country in East-Asia, constituting the southern part of the Korean Peninsula. Its total area is approximately 100,000 km² and 66% forest land and 19% agricultural land. In the humid temperate

climatic condition, national mean annual precipitation (MAP, mm) and temperature is 1,300 mm, 23°C. The basis of parent rock is the granite gneiss (32.4%), granite (22.3%), schist (10.3%), limestone (10.1%). Inceptisols is constituted 69.2 percent of the Soil taxonomic system and there are 7 classes of Soil Order (RDA, 2011). Jeju island, however, is constituted basalt of Andisols class (Fig. 1).

Fig. 1. Soil classification and digital elevation map (DEM) of South Korea.

Soil data and laboratory analysis The original soil data ($n = 5,023, 78,148$ layer) was collected by Korean Rural Development Administration (RDA) from 1968 to 1990. The data includes chemical and physical properties such as soil texture, organic matter, and a limited number of gravel and bulk density contents. So we match the Korean soil map (RDA)'s location using the soil series. Finally, we selected approximately 78,148 soil layer data and used for Digital soil mapping. The location of soil sample based on soil profiles in "Taxonomical Classification of Korean Soils" (RDA, 2011). The organic matter analyzed by Turine method and multiply using 1.724 conversion factor to calculate soil organic carbon. Bulk density was calculated the mass of an dried sample of soil per unit bulk volume.

Estimation of bulk density According to before study (Hong et al., 2013), we predict mineral bulk density (BD_{min}) using pedotransfer functions (PTFs) from sand content and depth based on the limited data in the database (Tranter et al., 2007).

$$BD_{min} = 1.017 + 0.0032 * sand + 0.054 * \log(mean \ depth) \quad (Eq. 2)$$

The influence of organic matter on bulk density is then calculated based on Adams (1973)' model.

$$BD \text{ (g cm}^{-3}\text{)} = \frac{100}{\frac{OM}{BD_{OM}} + \frac{100 - OM}{BD_{\min}}} \quad (\text{Eq. 3})$$

where, organic matter is the mass percentage of organic matter (g 100 g⁻¹), and BD_{OM} = average organic matter bulk density = 0.224 g cm⁻³, and BD_{min} is the mineral bulk density g cm⁻¹).

Because of the different mineralogical composition of andisols a separate model that predicted BD from OM content was derived from the Tempel et al. (1996).

$$BD \text{ (g cm}^{-3}\text{)} = 1.02 - 0.156 \log(OM) \quad (\text{Eq. 4})$$

Calculation of soil organic carbon density and carbon stocks Although SOC density and SOC stock are often used interchangeably in literature (Minasny et al., 2006), they differ in scale and unit. SOC density is the SOC mass per unit area for a given depth which can be calculated as:

$$SOC \text{ density (ton C ha}^{-1}\text{)} = \sum_{i=1}^k \frac{B_i C_i D_i}{10} (1 - G_i) \quad (\text{Eq. 5})$$

where, B_i is the Bulk density (g cm⁻³) and, C_i is Carbon concentration (g kg⁻¹), D_i is the thickness (cm) and G_i is the gravel and stone content (%).

SOC stock is the actual SOC mass for a given soil depth and area. It was calculated by summing up the product of mean SOC density per grid and the number of grid.

$$SOC \text{ stock (Mt)} = \sum_{i=1}^n \{SOC_{di} * A_i\} * 10^{-3} \quad (\text{Eq. 6})$$

where, SOC_{di} is SOC density (ton C ha⁻¹), A_i is area of grid cell and 10⁻³ is conversion factor.

Environmental covariates Environmental variables were selected to represent the key soil-forming factors of climate, parent material, relief and age based on the scorpan model (McBratney et al., 2003). Temperature affects the accumulation rate of carbon in soil, and temperature variables are widely used in prediction of SOC (Jobbágy and Jackson, 2000). Mean annual temperature (MAT) and mean annual precipitation (MAP) over thirty years (1970 - 2000) of the study area was obtained from Korea Meteorological Administration. These climate data collected from 67 sampling site in study area and reanalyzed and resampling to make 30 × 30 m grid layer.

Topography dataset came from the digital elevation model provided by the National Geographic Information Institute (<http://air.ngii.go.kr/gis/gisSrchInfoView.do>). Slope gradient, Analytical Hillshading, Topographic wet-

ness index (TWI), Terrain ruggedness index (TRI), Topographic openness, and Mid slope position (MSP) multi-resolution index were derived from the DEM using the System for Automated Geoscientific Analysis (SAGA) software (<http://www.saga-gis.org/en/index.html>). Soil texture divide by 7 classes such as Sandy, Coarse_loamy, Coarse_silty, Fine_loamy, Fine_silty, Clay. Soil parent divided by the Acidic (igneous_rocks), Neutral (igneous_rocks), Basic (igneous_rocks), Volcanic_ash, Sedentary rocks, Tertiary, Quaternary, Metamorphic_rocks, extra. Relationships among SOC and BD with the environmental variables are presented in Table 1.

Table 1. Description of environmental covariate.

Type	Name (code name)	Unit	Variables	Mean	Min	Max
Topographic data	Elevation (DEM)	m	cont. [†]	244	0	1,857
	Slope (slope)	%	cont.	16	0	86
	Analytical hillshading (hillshading)	radian	cont.	1.01	0.02	2.32
	Topographic wetness index (TWI)	unitless	cont.	5	-2	12
	Terrain ruggedness index (TRI)		cont.	7.28	0	1,302
	Topographic openness - positive (TOP)		cont.	1.38	0.03	2.28
	Topographic openness - negative (TON)		cont.	1.40	0.33	1.76
	Mid-slope position (MSP)	deg	cont.	27.6	0	55.8
Climate data	Mean annual temperature (MAT)	°C	cont.	11	3	17
	Mean annual precipitation (MAP)	mm	cont.	114	69	310
Soil map	Texture (texture)		cat. [‡]	7 classes		
	Parent (parent)		cat.	9 classes		

[†]cont.: continuous variable, [‡]cat.: categorical variable.

Equal-area spline profile function The spline function is based on the quadratic spline model of Bishop et al. (1999). This function was fitted to the profile values of the target soil organic carbon and bulk density using the es-splineTool of ithir package from R software to convert to horizon-based value to the IPCC standard depth (0 - 30 cm). The smoothing parameter lambda chosen was 0.1 (Malone et al., 2009).

Spatial predict model and validation To predict spatial distribution of 0 - 30 cm SOC stock in South Korea, we applied Cubist predict model. The Cubist model is currently a very popular model. Its popularity is due to its ability to ‘mine’ non-linear relationships in data (Malone et al., 2017). This model is based on the M5 algorithm of Quinlan (1987), and is implemented in the R cubist package.

In order to calibrate SOC predict model, the splined data were randomly split into 75% for training the model, and 25% data for validation. For validation the model, validated with 30% of the data using 3 statistical parameters: R^2 , MAE, RMSE. The R^2 is the coefficient of determination of linear regression (Eq. 7), between the observed values and predicted values. MAE is the mean error the absolute prediction error (MAE) and RMSE correspond to root mean square error (Eq. 8, Eq. 9).

$$R^2 = \frac{\sum_{i=1}^n (pred_i - \overline{obs})^2}{\sum_{i=1}^n (obs_i - \overline{obs})^2} \quad (\text{Eq. 7})$$

$$MAE = \frac{1}{n} \sum_{i=1}^n (obs_i - pred_i) \quad (\text{Eq. 8})$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (obs_i - pred_i)^2} \quad (\text{Eq. 9})$$

where, $pred_i$ and obs_i are predicted and observed SOC stock at site i ; n is the number of samples; R is the Pearson correlation coefficient between the observed and predicted values.

For the uncertainty we used bootstrap model prediction on the validation data. We estimate the confidence interval of those prediction for each point.

Results

descriptive statistics and estimated bulk density Equal-area splines (R package) modeled the depth-wise distribution and generated a continuous SOC and BD profile 0 - 100 cm depth. SOC content was highly variable and ranged from 5 to 226 ton C ha⁻¹ for the 0 - 30 cm and from 11 to 917 ton C ha⁻¹ in the 0 - 100 cm (Table 2). The mean SOC at 0 - 30 cm was 46 ton C ha⁻¹ and that for the 0 - 100 cm was 102 ton C ha⁻¹. The mean SOC for the 0 - 30 cm is almost half of 0 - 100 cm depth. Also BD was on average at a depth is 1.15 g cm⁻³ and 1.22 g cm⁻³ respectively.

Because of mineralogical difference, we used two equation for estimate bulk density (Eq. 2, Eq. 3).

Table 2. Descriptive statistics of soil organic carbon stock (ton C ha⁻¹) and bulk density (g cm⁻³).

Parameter	Depth	Mean	SD	Min	Max	Range
SOC stock	0 - 30	46	41	5	226	221
BD		1.15	0.19	0.56	1.45	0.88
SOC stock	0 - 100	102	109	11	917	906
BD		1.22	0.18	0.57	1.51	0.94

The boxplot of estimated Bulk density shown Fig. 2. In both soil type inland soil (non-volcanic) and island soil (volcanic), the estimated Bulk density value increase with the depth of the soil.

Fig. 2. The boxplot of estimated bulk density: non-volcanic and volcanic soil.

Importance of covariate in predicting SOC stock As shown in Table 3, the Cubist, based on the whole data set (78,148 layer data), used Texture, Parent, DEM and PRCP as conditions to perform the Cubist for SOC stock for 0 - 30 cm depth. However, TRI, PRCP, DEM, TOP, TON, TWI, MSP, Tavg, TON, DAH, Hillshading and Valley Depth were used as environmental covariates to predict the SOC stock. Among the environmental covariates, MAP, DEM, TWI, TRI, TOP, Slope and PRCP showed more influence on SOC and their spatial distributions.

Table 3. Usage (%) of covariates in the Cubist model for predicting SOC.

Attribute usage		SOC stock (0 - 30 cm depth)		
Conditions	Texture (93.9%)	Parent (78.5%)	DEM (67.3%)	MAP (29%)
Environmental covariates used in Cubist model	MAP (98.9%)	DEM (95.7%)	TRI (95.7%)	TOP (94.6%)
	TWI (91.8%)	Slope (87.9%)	TON (87.9%)	MAT (85.2%)
	MSP (83.8%)	Hillshading (25.3%)	DAH (12.5%)	
Attribute		SOC stock (0 - 100 cm depth)		
Conditions	Parent (90.4%)	Texture (50.0%)	DEM (43.9%)	TON (24.0%)
	MAT (11.6%)	MAP (10.8%)	TRI (3.8%)	
Environmental covariates used in Cubist model	MAP (89.5%)	TON (84.6%)	DEM (79.5%)	TOP (73.5%)
	MAT (68.7%)	TRI (66.6%)	Slope (65.9%)	TWI (57.8%)
	MSP (50.5%)	Hillshading (22.5%)	ValleyDepth (10.4%)	DAH (9.3%)

Validation of Cubist model prediction The overall cubist prediction for both depths (0 - 30 and 0 - 100 cm) is good with mean R^2 of 0.78 and 0.62, respectively (Table 4). When compared between two depths the cubist did better in predicting 0 - 30 cm than 0 - 100 cm SOC stock.

Table 4. Validation results of Cubist in predicting SOC in 0 - 30 cm and 0 - 100 cm depth.

Repeat	R ²	Concordance	MSE	RMSE	R ²	Concordance	MSE	RMSE
C stock (ton C ha ⁻¹) 0 - 30 cm					C stock (ton C ha ⁻¹) 0 - 100 cm			
1	0.78	0.88	382	19.55	0.62	0.75	4,743	68.87
2	0.77	0.88	387	19.68	0.62	0.76	4,653	68.21
3	0.79	0.88	369	19.23	0.62	0.74	4,947	70.34
4	0.78	0.88	369	19.23	0.63	0.75	4,718	68.69
5	0.78	0.88	378	19.46	0.62	0.74	4,850	69.64
6	0.77	0.88	381	19.54	0.62	0.76	4,771	69.08
7	0.78	0.88	371	19.28	0.62	0.76	4,729	68.77
8	0.78	0.88	391	19.81	0.64	0.77	4,474	66.89
9	0.77	0.87	392	19.82	0.63	0.76	4,567	67.58
10	0.78	0.88	375	19.37	0.62	0.75	4,766	69.04
Mean	0.78	0.88	380	19.50	0.62	0.75	4,722	68.71

The performance of cubist SOC stock model, R² ranged so stability from 0.77 to 0.79 for 0 - 30 cm and 0.62 - 0.64 for 0 - 100 cm. The variation of R² for SOC stock from cubist model, it is due to random selection effect of calibration and validation datasets.

The mean RMSE for 0 - 30 cm SOC stock and 0 - 100 cm SOC stock was 19.5 and 68.7, respectively. The 0 - 100 cm SOC stock RMSE higher than 0 - 30 cm SOC stock.

SOC stock The average SOC stocks calculated for two soil depths, it was 35.2 ton C ha⁻¹ for 0 - 30 cm and 87.1 ton C ha⁻¹ for 0 - 100 cm depth (Table 5). In the SOC stock map, for top 30 cm soil depth, most of the country area had SOC stocks below 40 ton C ha⁻¹ and some of mountain areas and Jeju island parts of the country have more than 50 ton C ha⁻¹ (Fig. 3). The total SOC stock (top 1 m soil depth) map also showed similar distribution to top 30 cm SOC map, but the overall SOC stock was high.

Table 5. Predicted mean SOC stock (ton C ha⁻¹) and confidence interval (95%).

	0 - 30 cm				0 - 100 cm			
	Mean	Min	Max	Range	Mean	Min	Max	Range
Cubist predicted	35.2	5	241	236	87.1	14.8	839.7	824.9
Lower limit	31.7	3.5	222.9	219.4	33.1	2.4	143.8	141.4
Upper limit	38.6	7.25	260.4	253.2	148.2	34.7	2,124.2	2,089.5
Prediction range	6.9	2.72	55.5	52.8	9.68	9.7	318.9	309.2
SD	2.1	0.8	16.9	16.1	2.94	1.15	96.9	95.8
SE	3.4	1.4	27.8	26.4	4.84	1.89	159.4	157.5

Fig. 3. Uncertainty (95%) and spatial distribution map of predicted SOC stocks (ton C ha^{-1}) in South Korea.

Conclusions

Digital Soil Mapping (DSM) can be defined as the creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from soil observation and knowledge and from related environmental variables (Lagacherie and McBratney, 2006). This study predicted the spatial distribution of SOC stock at two soil depth intervals (0 - 30, 0 - 100 cm) and quantified its stock (ton C ha^{-1}) for national scale. The mean estimated SOC stock at 0 - 30 cm soil depth was $35.2 \text{ ton C ha}^{-1}$ and for 0 - 100 was $87.1 \text{ ton C ha}^{-1}$. The SOC stock map could be used for future soil organic carbon stock and predicting potential SOC sequestration.

Acknowledgement

This study was carried out with the support from project (PJ015102) of Rural Development Administration, Republic of Korea.

References

- Adams, W.A. 1973. The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. *J. Soil Sci.* 24(1):10-17.
- Batjes, N.H. 1996. Total carbon and nitrogen in the soils of the world. *Eur. J. Soil Sci.* 47(2):151-163.
- Bishop, T.F.A., A.B. McBratney, and G.M. Laslett. 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* 91(1):27-45.
- Hong, S.Y., B. Minasny, K.H. Han, Y. Kim, and K. Lee. 2013. Predicting and mapping soil available water capacity in Korea. *PeerJ* 1(1):e71.
- Jobbágy, E.G. and R.B. Jackson. 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* 10(2):423-436.
- Jobbágy, E.G. and R.B. Jackson. 2001. The distribution of soil nutrients with depth: Global patterns and the imprint of plants. *Biogeochemistry* 53(1):51-77.
- Lagacherie, P. and A.B. McBratney. 2006. Chapter 1 Spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping. p. 3-22. In P. Lagacherie, A.B. McBratney, and M. Voltz (eds.) *Developments in Soil Science*, Vol. 31.
- Malone, B.P., A.B. McBratney, B. Minasny, and G.M. Laslett. 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* 154(1):138-152.
- Malone, B.P., Q. Styc, B. Minasny, and A.B. McBratney. 2017. Digital soil mapping of soil carbon at the farm scale: A spatial downscaling approach in consideration of measured and uncertain data. *Geoderma* 290:91-99.
- McBratney, A.B., M.L. Mendonça Santos, and B. Minasny. 2003. On digital soil mapping. *Geoderma* 117(1):3-52.
- Minasny, B., A.B. McBratney, M.L. Mendonça-Santos, I.O.A. Odeh, and B. Guyon. 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Aust. J. Soil Res.* 44(3):233-244.
- Powers, J.S. and W.H. Schlesinger. 2002. Relationships among soil carbon distributions and biophysical factors at nested spatial scales in rain forests of northeastern Costa Rica. *Geoderma* 109(3):165-190.
- Quinlan, J.R. 1987. Decision trees as probabilistic classifiers. p. 31-37. In P. Langley (ed.) *Proceedings of the Fourth International Workshop on Machine Learning*. University of California, Irvine.
- RDA (Rural Development Administration). 2011. Taxonomical classification of Korean soils. RDA, Suwon, Korea.
- Tempel, P., N.H. Batjes, and V.W.P. van Engelen. 1996. IGBP-DIS soil data set for pedotransfer function development, Wageningen.
- Tranter, G., B. Minasny, A.B. Mcbratney, B. Murphy, N.J. McKenzie, M. Grundy, and D. Brough. 2007. Building and testing conceptual and empirical models for predicting soil bulk density. *Soil Use Manage.* 23(4):437-443.
- Yang, Y., J. Fang, P. Smith, Y. Tang, A. Chen, C. Ji, H. Hu, S. Rao, K. Tan, and J.-S. He. 2009. Changes in topsoil carbon stock in the Tibetan grasslands between the 1980s and 2004. *Global Change Biol.* 15(11):2723-2729.