1 **Using an inexpensive color sensor for rapid assessment of soil**
2 **organic carbon**

3 Roxanne Stiglitz^a, Elena Mikhailova^{a,*}, Christopher Post^a, Mark
4 Schlautman^b, Julia L. Sharp^c

5

6 ^a *Department of Forestry and Environmental Conservation, Clemson University,*
7 *Clemson, SC, 29634, USA*

8 ^b *Department of Environmental Engineering and Earth Sciences, Clemson University,*
9 *Anderson, SC, 29625, USA*

10 ^c *Department of Mathematical Sciences, Clemson University, Clemson, SC, 29634, USA*

11

12

13 Corresponding author. Current address: Department of Forestry and Environmental
14 Conservation, Clemson University, 261 Lehotsky Hall, Clemson, SC, 29634, USA. Tel.
15 864-656-3535, Fax: 864-656-3304.

16 Email address: cleanam@clemson.edu (E.A. Mikhailova).

17

ABSTRACT

Quantifying soil organic carbon (SOC) is important for soil management, precision agriculture, soil mapping and carbon dynamics research. Inexpensive sensor technologies offer the potential for rapid quantification of SOC in laboratory samples as well as in the field. The objective of this study was to use a commercially-available color sensor to develop SOC prediction models for both dry and moist soils from the Piedmont region of South Carolina. Thirty-one soil samples were analyzed for lightness to darkness, redness to greenness, and yellowness to blueness (CIEL*a*b*) color using a Nix ProTM color sensor. Soil color was measured under both dry and moist soil conditions and the depth of each soil sample was also recorded. Using L*, a*, b* and soil depth for each sample as initial predictors, regression analyses were conducted to develop SOC prediction models for dry and moist soils. The resulting residual plots, root mean squared errors (RMSE), and coefficients of determination (R^2) were used to assess model fits for predicting the SOC content of soil. Cross validation was conducted to determine the efficiency of the predictive models and the mean squared prediction error (MSPE) was calculated. The final models included soil depth, L*, and a* as independent variables (dry soils $R^2 = 0.7978$ and MSPE = 0.0819, moist soils $R^2 = 0.7254$ and MSPE = 0.1536). The results suggest that soil color sensors have potential for rapid SOC determination, and soil depth and color are useful in predicting SOC content in soils.

Keywords: Carbon model, CIEL*a*b*, Nix ProTM, Regression analysis, Soil color coordinates, Ultisols

1. Introduction

Soil organic carbon (SOC) is one of the key soil properties related to ecosystem services (Adhikari and Hartemink, 2016) and it is measured by the Natural Resource Conservation Services (NRCS) to determine soil quality (Karlen et al., 2003, USDA/NRCS, 1996). Organic carbon in soil serves many purposes in soil fertility and structure by improving water retention and infiltration, improving soil texture, promoting soil organism growth, and by holding essential nutrients in the soil for healthy plant growth and production (Oades, 1984, Fontaine et al., 2003). In addition, soils play a major role in the carbon cycle by sequestering carbon dioxide from the atmosphere which would otherwise add to the effects of climate change (Li et al., 2006, Kheir et al., 2010). Disturbances in normal soil environments, such as deforestation and thawing permafrost, can lead to excessive release of stored carbon in the form of greenhouse gasses, such as carbon dioxide and methane, into the atmosphere (Potter, 1999; Christensen et al., 2004). Adhikari and Hartemink (2016) argue that soil is a vast determinant of a nation's economic standing and is linked to ecosystem service. Given the importance of soil and SOC both globally and agriculturally, there is a need for methods of rapid soil analysis and SOC determination that are inexpensive and easy-to-use.

It is well known that SOC content influences the color of a soil (Baumgardner et al., 1969). Studies have shown that because of this, it is possible to use a soil's reflectance to determine SOC content, making it possible to develop prediction models based on soil color (Bartholomeus et al., 2008). For this reason, many have turned to using visible near-infrared spectroscopy to determine SOC content in soils (Morgan et al.,

2009, Vasques et al., 2008). However, spectrometers can be expensive and many soil scientists may not be familiar with the resulting spectra data.

In a study by Wills et al. (2007), soil color in Munsell Color Chart notation, along with other soil qualities, were used to create a SOC prediction model for agricultural and prairie soils of the Midwest United States. Soil color value and chroma along with depth of the soil sample produced the most accurate SOC prediction model. However, SOC predictions can be limited based on regional soils. Different soil types will appear different in color based on SOC origin and soil mineralogy making it necessary to gather a large soil sample set that encompasses all soil types and SOC content before a universal SOC prediction model can be developed (Bartholomeus et al., 2008). Studies have shown that there is significant variation among Munsell Color Charts that can result in inaccurate color measurements which would lead to inaccurate SOC predictions (Sanchez-Maranon et al., 2005). In addition, Munsell Color Chart notation does not allow for simple statistical analysis which could complicate the process of creating a SOC prediction model for various regional soils (Kirillova et al., 2014).

Fortunately, there are a number of color systems to classify the color of soils that can be used in soil science (Viscarra Rossel et al., 2006). Recently, an inexpensive color sensor (Nix ProTM) was evaluated for its ability to determine soil color (Stiglitz et al., 2016a). The Nix ProTM produces color results in lightness to darkness, redness to greenness, and yellowness to blueness (CIEL*a*b* notation) and other color systems, is rechargeable and portable, and has its own light source making it a great mobile alternative to the Munsell Color Chart. The Nix ProTM offers a new method of color analysis that is accurate, rapid, and convenient for statistical analysis (Stiglitz et al.,

2016a,b). Using the Nix ProTM as a colorimeter would assist in gathering data necessary for developing SOC prediction models efficiently and reliably. The objectives of this study were (i) to gather soils data from the Piedmont region of South Carolina for analysis, (ii) create a SOC prediction model for dry soils of the Piedmont region of South Carolina, and (iii) create a SOC prediction model for moist soils of the Piedmont region of South Carolina.

2. Materials and methods

2.1 Study area and soil samples

The study area and samples for this experiment are as described previously by Stiglitz et al. (2016a) and were collected from the Piedmont region of South Carolina. For development of the predictive models, thirty-one samples (i.e., training set) were gathered from thirteen soil pits at the Simpson Agricultural Experimental Station (http://www.clemson.edu/public/researchfarms/beef_cattle/) near Pendleton, South Carolina. Three soil orders were represented in the collected samples (Fig. 1): Cecil clay loam (Cecil fine, kaolinitic, thermic Typic Kanhapludults), Pacolet sandy loam (Pacolet fine, kaolinitic, thermic Typic Kanhapludults), Cartecay-Chewacla complex (Cartecay-Chewacla complex of coarse-loamy, mixed, semiactive, nonacid, Thermic Aquic Udifluvents and fine-loamy, mixed, active, Thermic Fluvaquentic Dystrudepts), Hiwassee sandy loam (Hiwassee fine-loamy, mixed, active, Thermic Fluvaquentic Dystrudepts), and Cecil sandy loam ((Cecil fine, kaolinitic, thermic Typic Kanhapludults) (Soil Survey Staff, 2016). The soil series that were collected are typical of the Blue Ridge Mountains, Piedmont, and Valley and Ridge regions of the eastern United States,

spanning from Virginia to Georgia, north to south, and from the coast to Alabama, Tennessee, and Kentucky. In addition, thirty-one separate samples were taken from the soil pits for the purpose of cross validation (i.e., validation set).

The depth for each soil sample collected was recorded. Subsamples of each soil were sent to the Clemson University Agricultural Service Lab (www.clemson.edu/public/regulatory/ag_svc_lab/index.html) for nutrient analysis and to the University of Georgia Soil, Plant and Water Analysis Lab (aesi.ces.uga.edu) to be analyzed for texture and total carbon content (Fig. 2 and Table 1). Samples were analyzed for texture using the standard NRCS soil textural triangle and soil carbon percent was determined by lost on ignition. Soil samples were oven dried, crumbled, and passed through a 2-mm sieve before being analyzed for color. Soil samples, about one inch in diameter, were placed on a plate and the surfaces of each sample were leveled to allow for the sensor to rest on a flat sample surface which prevented any outside light from entering the viewing area of the sensor. Dried soil samples were moistened using a water dropper to dampen the soil surface. The soil samples were then analyzed for color using a Nix ProTM color sensor for both moist and dry soil conditions with results recorded in CIEL*a*b* following the methods described previously by Stiglitz et al. (2016a). The Nix ProTM color sensor cost \$349 and is controlled via Bluetooth using a free to download mobile application. The sensor has its own LED (light emitting diode) light source, rechargeable battery, and produces color results in various color systems such as CIEL*a*b*, RGB (red, green, blue), and XYZ (red, green, blue).

2.2 Development of SOC prediction models

Regression analysis was conducted using SAS Studio software (SAS Institute Inc., 2014) for the thirty-one soil samples (training set) using measured values from the soil sampling, nutritive analysis and color determination. The dependent variable in all regression models was the soil sample SOC%, and specific predictor variables considered were depth of the soil sample, and CIEL*a*b* color coordinates to keep the model as simple as possible. Initially, all chosen predictors (depth, L*, a*, and b*) were included in the model to determine which were useful in predicting SOC content of the soil. Using the coefficient of determination (R^2), root mean squared error (RMSE), and residual plots, model fit was assessed. Predictors that were determined insignificant were removed and the model was run again. This process was repeated until only significant predictors remained in the carbon prediction model. A level of significance of 0.05 was used for all regression analyses. Once the models were constructed, the predicted SOC content of the dry and moist soil samples was found using the selected soil sample predictors. The actual SOC content was then plotted against the predicted SOC content for comparison (Fig. 3).

2.3 Cross validation of the selected SOC prediction models

Thirty-one additional soil samples (validation set) were gathered from the same soil pits from Simpson Agricultural Station for the purpose of cross validation and were not utilized for model development. Depth and Nix ProTM CIEL*a*b* color coordinates were recorded for each sample under both dry and moist soil conditions. The data for the

samples were used to predict the SOC content using the two prediction models for dry and moist soils. The actual and predicted SOC content of the samples were compared and the mean squared prediction error (MSPE) was calculated for both dry and moist SOC prediction models to determine how well the models would predict SOC content of soils. The smaller the MSPE, the better the model would be at predicting SOC.

3. Results and discussion

3.1 Development of the SOC prediction model for dry soils

A multivariate regression model was conducted using sample depth, L^* , a^* , and b^* as parameters for predicting SOC% for dry soil samples (Table 2). Table 3 shows that b^* is linearly correlated with SOC for moist and dry soils, and is also linearly correlated with sample depth and with a^* for dry soils, suggesting that it may not be a desirable predictive parameter to include in the model because of its correlation with the other predictors (i.e., there may be issues with multicollinearity). When all of the predictors were included in the model predicting SOC%, the residual plots indicated that a quadratic b^* effect may be necessary to include in the model. In addition, in the dry soil samples model with all of the predictors, b^* was not a significant predictor of SOC%, adjusting for the other predictors in the model ($p=0.1107$; Table 4). After considering the quadratic b^* effect in the model, the RMSE and R^2 were only marginally improved and were thus left out of the final model. For the dry soil final model, p-values indicate that sample depth ($p\text{-value} = 0.0011$), L^* ($p\text{-value} < 0.0001$), and a^* ($p\text{-value} = 0.0002$) were useful in predicting SOC (RMSE=0.42490 and $R^2 = 0.7978$; Table 5). Past studies have shown that there is a correlation between SOC and lightness and darkness of a soil (Shields et al.,

1968). This could explain why L^* had the most significant p-value for dry soils. In addition, it has been reported that a close correlation exists between b^* and the iron oxide content of soils (Schwertmann, 1993; Scheinost and Schwertmann, 1995), which at least partially explains the correlation observed here between b^* and sample depth. Iron content tends to increase as the degree of weathering increases for a soil horizon, thus, iron content tends to increase with soil depth (Rebertus and Buol, 1985). More importantly, it suggests that including b^* in a predictive model for SOC% would not be desirable for high iron content soils such as the ones tested in this study. Soil texture was not considered for the models, but soil surface texture and particle size may affect reflectance and result in differing color results which would affect SOC prediction models. Figure 3a shows a plot of actual SOC% versus predicted SOC% for the final predictive model for dry soils.

3.2 Development of the SOC prediction model for moist soils

Following the general procedures described above for the dry soils, multivariate regression analysis was utilized to develop a predictive model for SOC% in moist soils considering initially sample depth, L^* , a^* , and b^* as predictive parameters (Table 2). Consistent with the dry soils, once again b^* was found to be linearly correlated with SOC but also again with sample depth, L^* and a^* for the moist soil samples (Table 3). In the model for moist soil samples that included all of the parameters, b^* was not a significant predictor of SOC% (p-value = 0.7353; Table 4). A quadratic b^* effect in the model was also considered, but the marginal improvement in predictive capabilities was not statistically significant. For the moist soil final model, the sample depth (p-value =

0.0020), L^* (p-value = 0.0043), and a^* (p-value = 0.0009) were useful in predicting SOC (RMSE = 0.49509 and $R^2 = 0.7525$; Table 5). Typically, SOC content decreases as soil depth increases (Jobbagy and Jackson, 2000). For this reason, depth was considered for both dry and moist soil prediction models. While the samples were moistened, the exact amount of moisture was not measured and, therefore, it is not possible to know exactly how soil moisture would affect the prediction of SOC content in soils. However, the resulting RMSE and R^2 suggests that the chosen parameters were still sufficient at predicting SOC content. Figure 3b shows a plot of actual SOC% versus predicted SOC% for the final model for moist soils.

3.3 Cross validation of the SOC prediction models

The R^2 values for dry and moist soil predictive models were comparable to those found by Wills et al. (2007) as all values were significant and accounted for a large amount of error within the model. A cross validation was conducted using the SOC predictive models for dry and moist soils and thirty-one additional soil samples (validation set) from the thirteen experimental soil pits. The models were used to predict the SOC content of the additional samples using their depth and color data. The actual and predicted SOC content of each additional sample was compared and the MSPE was calculated for the dry soils predictive model (MSPE = 0.0819; Table 5) and moist soils predictive model (MSPE = 0.1536; Table 5). The results suggest that the models are sufficient at predicting SOC content. Having models that can predict SOC content of soils, and thereby soil quality, would be a useful addition to the Soil Quality Test Kit Guide (http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1044790.pdf)

which aims determine a soil's "ability to perform basic functions" by gathering minimum data in field settings.

The results of the regression analyses suggest that sample depth, L^* , and a^* are the best variables for predicting SOC in both dry and moist South Carolina Ultisols. There was a notable trend in SOC content compared to depth, specifically that SOC content decreased as depth increased as seen in Table 1 making depth a wise choice to include in a prediction model. The L^* value of each sample also seemed to have a notable impact of the SOC content. Given that Ultisols are red in color, it would be expected that the a^* value for redness would influence soil characteristics such as SOC content. While b^* did not prove to be a strong predictor of SOC content in the soils of the Piedmont region of South Carolina, it may be a sufficient indicator of SOC in soils of another type that contain either yellow or blue pigments. The R^2 values for dry and moist soil predictive models were respectable, suggesting that the models created are sufficient for predicting SOC content of soils of South Carolina. It is important to note that the created models may also be sufficient predictors of SOC for soils of the same type that are typical of the Blue Ridge, Piedmont, and Valley and Ridge regions of the United States. The MSPE values for the two models suggests that while both models are sufficient predictors of SOC content, the model for dry soils is better at predicting SOC.

4. Conclusions

The models developed in this study likely have limitations. The soil samples used for this experiment were from an agricultural farm in the Piedmont region of South Carolina and were low in SOC content and high in iron content which may negatively

247 impact the ability to predict SOC based on color. Soils that are not used agriculturally
248 may have different characteristics that would impact SOC content and the accuracy of the
249 models may vary. It is likely that different models will need to be developed for soils of
250 different types, regions, and land-uses. This would, however, not require site-specific
251 models, as it should be possible to develop models that are applicable across large areas
252 with similar soil types and land uses. Furthermore, each method of color analysis would
253 potentially produce carbon models unique to the device being used, as different
254 illuminations, fields of view, and scanning angles may produce varying color results.

255 Regardless of the current limitations, there is promise in the Nix ProTM
256 technology and its ability to predict SOC content of soils. Carbon prediction models offer
257 a rapid and inexpensive method of soil quality assessment. Rapid SOC analysis could
258 assist in determining best management practices or soil reclamation methods that would
259 help preserve and restore farmland and other habitats. Prediction models would also offer
260 a means of monitoring carbon pools in areas being affected by deforestation, permafrost
261 thawing, and climate change (Potter, 1999; Christensen et al., 2004). There is notable
262 advantage in determining a close estimation of SOC without having to send samples to a
263 nutrient analysis laboratory, which takes time to produce results and can become costly
264 depending on the number of samples being analyzed. Rapid SOC field assessment tools,
265 such as this proposed soil sensor based method, can enable a much higher spatial density
266 of samples which may improve our understanding of the distribution of SOC across the
267 landscape. Many researchers in various fields of science would benefit from the
268 development of regional carbon predictions models that require only a few soil

269 parameters, such as sample depth and soil color, to function properly and provide an
270 estimate of SOC content.
271

272 **Acknowledgments**

273 Clemson University provided funding for this study; Technical Contribution No. 6434 of
274 the Clemson University Experiment Station. This material is based upon work supported
275 by NIFA/USDA, under projects: SC-1700452, and SC-1700462.

276

References

- Adhikari, K., Hartemink, A.E. 2016. Linking soils to ecosystem services – A global review. *Geoderma* 262, 101-111.
- Agricultural and Environmental Services Laboratories. 2011. University of Georgia. <<http://www.aesl.ces.uga.edu>>.
- Agricultural Service Laboratory. 2014. Clemson University. <http://www.clemson.edu/public/regulatory/ag_svc_lab/index.html>.
- Bartholomeus, H.M., Schaepman, M.E., Kooistra, L., Stevens, A., Hoogmoed, W.B., Spaargaren, O.S.P. 2008. Spectral reflectance based indices for soil organic carbon quantification. *Geoderma* 145, 28-36.
- Baumgardner, M.F., Kristof, S., Johannsen, C.J., Zachary, A. 1969. Effects of organic matter on the multispectral properties of soils. Purdue University. *Agr. Exp. Station J.* 3939, 413-422.
- Christensen, T.R., Johansson, T., Akerman, H.J, Mastepanov, M., Malmer, N., Friborg, T., Crill, P., Svensson, B.H. 2004. Thawing sub-arctic permafrost: Effects on vegetation and methane emissions. *Geophysical Research Letters*. 31 (4), 1-4.
- Fontaine, S., Mariotti, A., Abbadie, L. 2003. The priming effect of organic matter: a question of microbial competition? *Soil Biol. Biochem.* 35, 837-843.
- Jobbagy, E.G, Jackson, R.B. 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* 10 (2), 423-436.
- Karlen, D.L., Ditzler, C.A., Andrews, S.S. 2003. Soil quality: Why and how? *Geoderma* 114, 145-156.

- 299 Kheir, R.B., Greve, M.H., Bocher, P.K., Greve, M.B., Larsen, R., McCloy, K. 2010.
 300 Predictive mapping of soil organic carbon in wet cultivated lands using
 301 classification-tree based models: The case study of Denmark. *J. Environ. Manage.*
 302 91, 1150-1160.
- 303 Kirillova, N.P., Vodyanitskii, Yu. N., Sileva, T.M. 2014. Conversion of soil color
 304 parameters from the Munsell System to the CIE-L*a*b* System. *Eurasian Soil*
 305 *Sci.* 48 (5), 468-475.
- 306 Li, Z.P., Han, F.X., Su, Y., Zhang, T.L., Sun, B., Monts, D.L., Plodinec, M.J. 2007.
 307 Assessment of soil organic and carbonate carbon storage in China. *Geoderma*
 308 138, 119-126.
- 309 Morgan, C.L.S., Waiser, T.H., Brown, D.J., Hallmark, C.T. 2009. Simulated in situ
 310 characterization of soil organic and inorganic carbon with visible near-infrared
 311 diffuse reflectance spectroscopy. *Geoderma* 151, 249-256.
- 312 Oades, J.M. 1984. Soil organic matter and structural stability: mechanisms and
 313 implications for management. *Plant Soil.* 76, 319-337.
- 314 Potter, C.S. 1999. Terrestrial biomass and the effects of deforestation on the global
 315 carbon cycle. *Oxford Journals.* 49 (10), 769-778.
- 316 Rebertus, R.A., Buol, S.W. 1985. Iron distribution in a developmental sequence of soils
 317 from mica gneiss and schist. *Soil Sci. Soc. Am. J.* 49, 713-720.
- 318 Sanchez-Maranon, M., Huertas, R., Melgosa, M. 2005. Colour variation in standard soil-
 319 colour charts. *Aust. J. Soil Res.* 43, p. 827-837.
- 320 SAS Institute Inc. 2014. SAS® Studio 3.1: User's Guide. Cary, NC. SAS Institute Inc.

- 321 Schwertmann, U. 1993. Relations between iron oxides, soil color and soil formation. *In*
 322 Bigham, J.M. and Ciolkosz, E.J. (ed.) Soil Color. SSSA Spec. Publ. 31. Soil
 323 Science Society of America (SSSA), Madison, WI. p. 51-69.
- 324 Scheinost, A.C. and Schwertmann, U. 1995. Predicting phosphate adsorption-desorption
 325 in a soilscape. *Soil Sci. Soc. Am. J.* 59, p.1575-1580.
- 326 Series Extent Explorer. 2016. University of California Davis.
 327 <http://casoilresource.lawr.ucdavis.edu/see/>
- 328 Sheilds, J.A., Paul, E.A., Arnaud, R.J. St., Head, W.K. 1968. Spectrophotometry
 329 measurement of soil color and its relationship to moisture and organic matter.
 330 *Can. J. Soil. Sci.* 48 (3), p. 271-280.
- 331 Simpson Beef Cattle Farm. 2014. Clemson University.
 332 http://www.clemson.edu/public/researchfarms/beef_cattle/
- 333 Soil Survey Staff. 2016. Natural Resources Conservation Service, United States
 334 Department of Agriculture. Web Soil Survey.
 335 <<http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm>>
- 336 Soil Quality Test Kit Guide. 1999. Natural Resources Conservation Service, United
 337 States Department of Agriculture.
 338 <http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1044790.pdf>.
- 339 Stiglitz, R., Mikhailova, E., Post, C., Schlautman, M., Sharp, J. 2016a. Evaluation of an
 340 inexpensive sensor to measure soil color. *Comput. Electr. Agr.* 121, 141-148.
- 341 Stiglitz, R., Mikhailova, E., Post, C., Schlautman, M., Sharp, J. 2016b. Teaching soil
 342 color determination using an inexpensive color sensor. *Nat. Sci. Educ.* In Press.

- 343 USDA/NRCS. 1996. Soil Quality Indicators: Organic Matter.
344 <http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_053150.pdf>.
- 345 Vasques, G.M., Grunwald, S., Sickman, J.O. 2007. Comparison of multivariate methods
346 for inferential modeling of soil carbon using visible/near-infrared spectra.
347 Geoderma 146, 14-25.
- 348 Viscarra Rossel, R.A., Minasny, B., Roudier, P., McBratney, A.B. 2006. Colour space
349 models for soil science. Geoderma 133, 320-337.
- 350 Wills, S.A., Burras, C.L., Sandor, J.A. 2007. Prediction of soil organic carbon content
351 using field and laboratory measurements of soil color. Soil Sci. Soc. Am. J. 71 (2),
352 380-388.

353 **List of Figures**

354 **Fig. 1.** Map showing the extend of the soil series collected for analysis.

355

356 **Fig. 2.** Example of soil profile (out of 7 total soil profiles used in the study) for practice
357 soil pit 2 used during 2014 Southeast Regional Collegiate Soils Contest (October 5-9,
358 2014) (Stiglitz et al., 2016).

359

360 **Fig. 3.** Plots of actual SOC% content versus predicted SOC% content for a) dry and b)
361 moist soil samples.

362

363

Fig. 1. Map showing the extend of the soil series collected for analysis (<http://casoilresource.lawr.ucdavis.edu/see/>).

Fig. 2. Example of soil profile (out of 7 total soil profiles used in the study) for practice soil pit 2 used during 2014 Southeast Regional Collegiate Soils Contest (October 5-9, 2014) (Stiglitz et al., 2016)

a)

b)

Fig. 3. Plots of actual SOC% content versus predicted SOC% content for a) dry and b) moist soil samples for the training data set.

Table 1

Selected soil properties for practice soil pit 2 (Stiglitz et al., 2016).

Horizon	Lower depth (cm)	Texture	Sand	Silt	Clay	SOC	pH in water	BS (%)	CEC (meq/100g)	P	K	Ca	Mg	Zn	Mn	Cu	B	Na
			(%)							(mg/kg)								
Ap	11	SL	70	14	16	1.3	5.0	42	4.8	10.0	23	299	107	2.5	14	0.25	0.15	4.5
Bt1	28	SCL	58	14	28	0.44	5.6	39	3.9	1.5	14	207	100	0.5	4	0.30	0.10	5.5
Bt2	59	SC/SCL	52	12	36	0.36	5.6	35	3.3	1.0	9	186	77	0.4	3	0.20	0.10	5.0
Bt3	90+	SC/C	46	8	46	0.26	5.5	30	4.0	1.0	9	217	111	0.4	1	0.25	0.15	6.5

*SL = sandy loam, SCL = sandy clay loam, SC = sandy clay, C = clay

Table 2 Example of dry and moist soil variables for practice soil pit 2.

Horizon	SOC (%)	Depth (cm)	L*	a*	b*
<u>Dry Soils</u>					
Ap	1.3	11	52.16	12.41	24.7
Bt1	0.44	28	52.61	19.54	31.53
Bt2	0.36	59	53.16	17.68	30.75
Bt3	0.26	90	49.12	19.89	30.42
<u>Moist Soils</u>					
Ap	1.3	11	34.24	12.32	20.63
Bt1	0.44	28	38.84	21.06	27.91
Bt2	0.36	59	37.52	19.32	28.17
Bt3	0.26	90	35.41	22.29	28.67

Table 3

Pearson correlation (r) values among soil variables for dry and moist soils (p-values in parentheses).

Parameter	SOC	Depth	L*	a*	b*
<u>Dry soils</u>					
SOC	---	-0.69681 (<0.0001)	-0.29296 (0.1097)	-0.62770 (0.0002)	-0.84271 (<0.0001)
Depth		---	-0.05174 (0.7822)	0.56299 (0.0010)	0.66960 (<0.0001)
L*			---	-0.35450 (0.0504)	0.21120 (0.2541)
a*				---	0.67985 (<0.0001)
b*					---
<u>Moist soils</u>					
SOC	---	-0.69681 (<0.0001)	-0.34680 (0.0560)	-0.70689 (<0.0001)	-0.72698 (<0.0001)
Depth		---	0.04513 (0.8095)	0.57325 (0.0007)	0.57238 (0.0008)
L*			---	0.02802 (0.8811)	0.62670 (0.0002)
a*				---	0.59742 (0.0004)
b*					---

Table 4

Parameter estimates and ANOVA results for initial SOC prediction models.

Model	Parameter	Parameter estimate	Parameter p-value	Model p-value	Root MSE	R-squared
Dry soils	SOC	8.51924	<0.0001	<0.0001	0.41189	0.8170
	Depth	-0.00812	0.0172			
	L*	-0.07588	0.0034			
	a*	-0.07699	0.0172			
	b*	-0.06942	0.1107			
Moist soils	SOC	5.63852	<0.0001	<0.0001	0.50339	0.7266
	Depth	-0.01079	0.0109			
	L*	-0.05643	0.1285			
	a*	-0.09163	0.0106			
	b*	-0.01686	0.7353			

Table 5

Parameter estimates and ANOVA results for final SOC prediction models.

Model	Parameter	Parameter estimate	Parameter p-value	Model p-value	Root MSE	R-squared	MSPE
Dry soils	SOC	8.50860	<0.0001				
	Depth	-0.01060	0.0011				
	L*	-0.10138	<0.0001	<0.0001	0.42485	0.7978	0.0819
	a*	-0.11292	0.0002				
Moist soils	SOC	5.70287	<0.0001				
	Depth	-0.01146	0.0020				
	L*	-0.06625	0.0043	<0.0001	0.49509	0.7254	0.1536
	a*	-0.09830	0.0009				