Pedosphere **28**(6): 856–872, 2018
doi:10.1016/S1002-0160(18)60048-7
ISSN 1002-0160/CN 32-1315/P

Published by Elsevier B.V. and Science Press

PEDOSPHERE

www.elsevier.com/locate/pedosphere

Transferability and Scalability of Soil Total Carbon Prediction Models in Florida, USA

Sabine GRUNWALD[1,*], Congrong YU[1,2,3,4] and Xiong XIONG[1]

[1] *Department of Soil and Water Sciences, University of Florida, Gainesville FL 32611 (USA)*
[2] *State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098 (China)*
[3] *College of Hydrology and Water Resource, Hohai University, Nanjing 210098 (China)*
[4] *Agricultural and Biological Engineering Department, University of Florida, Gainesville FL 32611 (USA)*

(Received May 24, 2017; revised June 13, 2018)

ABSTRACT

The applicability, transferability, and scalability of visible/near-infrared (VNIR)-derived soil total carbon (TC) models are still poorly understood. The objectives of this study were to: i) compare models of three multivariate statistical methods, partial least squares regression (PLSR), support vector machine (SVM), and random forest methods, to predict soil logarithm-transformed TC (logTC) using five fields (local scale) and a pooled (regional-scale) VNIR spectral dataset (a total of 560 TC spectral datasets), ii) assess the model transferability among fields, and iii) evaluate their up- and downscaling behaviors in Florida, USA. The transferability and up- and downscaling of the models were limited by the following factors: i) the spectral data domain, ii) soil attribute domain, iii) methods that describe the internal model structure of VNIR-TC relationships, and iv) environmental domain space of attributes that control soil carbon dynamics. All soil logTC models showed excellent performance based on all three methods with $R^2 > 0.86$, bias $< 0.01\%$, root mean squared error (RMSE) $= 0.09\%$, residual predication deviation (RPD) $> 2.70\%$, and ratio of prediction error to interquartile range (RPIQ) > 4.54. The PLSR method performed substantially better than the SVM method to scale and transfer the TC models. This could be attributed to the tendency of SVM to overfit models, while the asset of the PLSR method was its robustness when the models were validated with independent datasets, transferred, and/or scaled. The upscaled soil TC models performed somewhat better in terms of model fit (R^2), RPD, and RPIQ, whereas the downscaled models showed less bias and smaller RMSE based on PLSR. We found no universal trend indicating which of the four limiting factors mentioned above had the most impact that constrained the transferability and scalability of the models. Given that several factors can impinge on the empirically derived soil spectral prediction models, as demonstrated by this study, more focus on their applicability and scalability is needed.

Key Words: attribute domain, multivariate statistical methods, partial least squares regression, random forest, support vector machine, soil carbon dynamics, visible-near infrared spectroscopy

Citation: Grunwald S, Yu C, Xiong X. 2018. Transferability and scalability of soil total carbon prediction models in Florida, USA. *Pedosphere*. **28**(6): 856–872.

INTRODUCTION

The assessment of soil total carbon (TC) across large land areas is critical to derive global and regional soil carbon budgets and better understand the interactions between carbon and other biogeochemical cycles. However, the cost and time involved in the measurements of TC with standard laboratory methods are impractical. Research has suggested that visible/near-infrared (VNIR) diffuse reflectance spectroscopy can provide robust and accurate estimations of TC and carbon fractions (McCarty *et al.*, 2002; Reeves III *et al.*, 2002; Viscarra Rossel *et al.*, 2006; Vasques *et al.*, 2009, 2010; Nocita *et al.*, 2011; Chaudhary *et al.*, 2012; McDowell *et al.*, 2012b). It is a fast, cheap, and non-destructive approach to characterize soil properties (Shepherd and Walsh, 2002). To promulgate the application of locally calibrated spectral soil carbon prediction models in other regions and at much larger spatial scales typically involves model transfer and/or scaling. The spectral soil carbon models are poised to contribute to spatially explicit regional and global carbon assessment and monitoring of soil carbon evolution. However, the uncertainties in regard to prediction quality across different soils and landscapes, transferability, and scalability of such models are still eminent (Grunwald *et al.*, 2011).

Scaling in ecology and earth sciences refers to the translation of information between or across spatial and temporal scales of organizational levels (Turner *et al.*, 1989; Blöschl and Sivapalan, 1995). 'Scale transformation' denotes the across-scale translation of informa-

*Corresponding author. E-mail: sabgru@ufl.edu.

tion through explicit mathematical expressions and statistical relationships (scaling equations) or process-based simulation (Blöschl and Sivapalan, 1995; Wu *et al.*, 2006). 'Scaling up' (or upscaling) translates information from finer scales (smaller grain sizes or extents) to broader scales (larger grain sizes or extents), whereas 'scaling down' (or downscaling) translates information from broader to finer scales (Blöschl and Sivapalan, 1995; Wu *et al.*, 2006). As the spatial scale increases from fine (field) to coarser scales (region, continent, and globe), the increasing extent and geographic domain space translates into increased variance of soil attributes (McBratney, 1998). Scale-independent behavior (*i.e.*, self-similar behavior) assumes that a coarser-scale system behaves like an average finer-scale system, implying that processes are linear. The nonlinear dynamics (*i.e.*, scale-dependent behavior) are still poorly investigated in the soil science discipline (Grunwald *et al.*, 2011). The distinction between scaling and transfer of soil models is critical because they entail inter- and extrapolations to varying degrees, causing uncertainty in model predictions.

Based on a broader definition, 'transfer' can be understood as applying a method/model derived from one set of empirical observations onto another set (or population). For instance, model transfer involves applying a calibrated VNIR-based soil prediction model to an independent validation spectral set to make predictions of soil carbon within a given geographic domain. 'Transfer' might also refer to the application of a soil model developed in one region (represented by a specific soil spectral dataset) to another region (*i.e.*, another spectral dataset) to predict soil properties of interest. In this case, the transfer of a model typically entails extrapolation, and its degree of extrapolation increases as the taxonomic distance of soils between the regions of 'model development' and 'model application' increases. Mallavan *et al.* (2010) asserted that the more similar the regions are in terms of soil environmental properties, the more likely it is to successfully transfer a soil prediction model. Several methods, such as the Gower similarity index (Gower, 1971) and soil taxonomic distance metrics (Minasny and McBratney, 2007), have been utilized to assess similarity among soil ecosystems.

A review of spatial scaling concepts and procedures used in digital soil mapping (DSM) was provided by Malone *et al.* (2013). Vasques *et al.* (2012a) assessed the scaling effects of soil carbon models, considering the geographic extent and grain size of models in Florida, USA. Vasques *et al.* (2012b) investigated the multiscale behavior of soil carbon and identified those environmental factors that imposed maximum effects on the predictions of soil carbon at three nested spatial scales. However, scaling studies of VNIR soil prediction models are rarely found in the literature. For example, Brown *et al.* (2005) found that about half of the transferred VNIR soil carbon models did not perform well in Montana, USA, despite pre-screening for spectral similarity. Minasny *et al.* (2009) transferred mid-infrared (MIR) spectral models that provided excellent performance to predict soil organic carbon (SOC) and TC, but were severely degraded when applied to other regions in eastern Australia. They asserted that the local calibration of spectral models is preferable to avoid the measurement biases between laboratories in different regions. In addition, they argued that the calibrated soil spectral functions are specific to the soil types in an area, limiting their transferability to other application sites. Spectral soil prediction models are abundantly found in the literature; however, knowledge gaps still exist in their transfer and up- and downscaling.

Shepherd and Walsh (2002) initiated the idea of 'ubiquitous' spectral libraries to predict soil properties. Once developed (calibration phase) and validated, these soil spectral libraries can be applied elsewhere, similar to pedotransfer functions (PTF). Brown *et al.* (2005) presented such an 'ubiquitous' global soil spectral library derived from VNIR spectra using a large soil dataset of 3 768 samples from the U.S. and additional 416 samples from 36 different countries in Africa, 104 from Asia, 75 from the Americas, and 112 from Europe. They obtained a validation root mean squared error (RMSE) of 7.9 g kg^{-1} for SOC that ranged from 0.0 to 536.8 g kg^{-1} with a median of 4.7 g kg^{-1}. They suggested that VNIR soil characterization has the potential to replace or augment the standard soil characterization techniques where rapid and inexpensive analysis is required. One limitation of the VNIR soil carbon modeling is that the predictions based on this modeling are relatively poor in the low-soil carbon attribute domain space (Vasques *et al.*, 2010; McDowell *et al.*, 2012b).

Numerous factors potentially influence the prediction performance of the transferability and scalability of soil spectral models. These factors include: i) number of samples used to build the spectral model, ii) range of observed soil carbon concentrations that is influenced by the different environmental factors that control the pedogenic processes to store or lose carbon in soils, iii) other soil properties (such as minerals, sesquioxides, and texture) that might mask, interfere, or amplify the identification of spectral predictors to infer on soil carbon, iv) differences in the measurement

protocols of soil carbon and spectral data, v) spectral instruments, vi) statistical methods used to develop soil carbon predictions, and vii) (dis)similarity between the soil samples used to develop the spectral prediction model and those that are used for estimation. Given the multitude of potential factors that might affect the application of VNIR soil carbon models to make predictions for unknown samples, the underlying motivation for this research was to design an experimental study to investigate the transfer and up- and downscaling behaviors of TC prediction models. The specific objectives were to i) compare the performance of three model types to predict soil TC using five fields (local scale) and a pooled (regional scale) VNIR spectral dataset, ii) assess the model transferability among five representative field sites in Florida, USA, iii) evaluate the upscaling behavior of TC prediction models from the local (field) to regional scale, iv) evaluate the downscaling behavior of TC prediction models from the regional to local scale, and v) examine the constraining factors in model transferability and scalability.

MATERIALS AND METHODS

Study area

Five fields (each approximately 0.25 km^2 in area) that represent prominent land use types in Florida were selected (see Xiong *et al.*, 2016). Table I provides a description of the main landscape characteristics of each field.

Field sampling

Soil samples were collected from each of the five

TABLE I

Characteristics of the five selected fields that represent prominent land use types in the study area in Florida, USA

Variable	Field 1	Field 2	Field 3	Field 4	Field 5
Sampling location	Ordway-Swisher Biological Station	San Felasco Hammock Preserve State Park	Econfina Creek Water Management Area	Santa Fe River Ranch	Myakka River State Park
Longitude	81°59′9″ W	82°27′31″ W	85°33′51″ W	82°29′40″ W	82°17′16″W
Latitude	29°41′23″ N	29°43′59″ N	30°26′42″N	29°55′45″ N	27°11′22″ N
Topography[a)]					
Elevation (m)	42.8	43.5	23.9	28.8	8.7
Slope (%)	1.2	1.2	2.9	2.4	0.2
Climate[b)]					
Maximum temperature (°C)	27.5	27.1	26.3	27.2	29.2
Minimum temperature (°C)	14.0	13.8	12.9	13.6	16.3
Precipitation (mm)	1 325	1 345	1 634	1 360	1 464
Surficial geology/ parent material[c)]	Cypresshead	Coosawhatchie	Citronelle	Coosawhatchie	Shelly sediments of Plio-Pleistocene
Land use[d)]	Xeric upland forest	Mesic upland forest	Pineland	Improved pasture	Rangeland
Vegetation					
NPP (kg C m^{-2})[e)]	7.91	13.60	9.07	7.50	8.13
NDVI[e)]	3.81	7.90	3.81	9.50	4.31
Dry biomass (kg m^{-2})[f)]	2.76	12.50	5.53	–	6.68
Soil[g)]					
Suborder	Psamment	Aquult, Psamment, Udept, Udult	Psamment	Udult	Aquod
AWHC (cm cm^{-1})	1.2	2.1	1.5	2.2	1.7
Clay (%)	1.2	5.2	3.7	4.6	1.9
Sand (%)	98.6	93.2	93.1	90.8	96.8

[a)] National Elevation Dataset (USGS, 1999).
[b)] Long-term maximum and minimum annual average temperature and long-term annual average precipitation between 1971 and 2000 from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) Climate Group.
[c)] USGS, 1999.
[d)] FFWCC, 2003.
[e)] Net primary productivity (NPP) and normalized difference vegetation index (NDVI) from the Moderate-Resolution Imaging Spectroradiometer (MODIS) for North American Carbon Project, 2005.
[f)] National Biomass and Carbon Dataset (NBCD), 2000.
[g)] Soil suborder, available water-holding capacity at 0–25 cm (AHWC), and clay content and sand content at 0–20 cm from the Soil Survey Geographic Database (SSURGO) (USDA-NRCS, 2009).

fields with the same unbalanced nested spatial sampling design (Lark, 2011) as described by Xiong *et al.* (2016). In each field, at first, nine main centers gridded at 200-m intervals were chosen to constitute the highest level of the hierarchy. Secondly, at each main center, one additional sampling point (sub-node) was collected 67 m away in a random direction. In similar pattern, the 2nd, 3rd, and 4th hierarchical sampling points were fixed at locations 22, 7, and 2 m away from their parent nodes, respectively. A total of 112 samples were collected at the depth of 0–20 cm in each field, totaling 560 samples within all five fields. The approximately 3-fold hierarchy has been proven to be effective in capturing soil variation and avoiding overlaps among different branches (Webster and Oliver, 2007).

Laboratory analysis

Soil TC content was measured using the dry combustion method in the laboratory. The soil cores were air dried in a drying room for five days at the temperature of 45–50 °C and then ground and passed through a 2-mm sieve. Approximately 50 g of each sample was ball milled for 3 min, from which approximately 500 mg was combusted at 900 °C for approximately 3 min on a Shimadzu TOC-5050 analyzer to measure the TC content. Miniscule amounts of inorganic carbon are found in Florida soils, and the SOC constitutes TC at many sites (Vasques *et al.*, 2010). Since soil moisture can impact the soil spectral reflectance values (Lobell and Asner, 2002), the ball-milled samples were oven dried at 40–45 °C for 12 h before scanning. After cooling for 1 h, the samples were scanned using the QualitySpec Pro spectroradiometer (Analytical Spectral Devices Inc., Boulder, USA) in the VNIR spectral range of 350–2 500 nm with a 1-nm interval spectral resolution. For each sample, four replicate scans were taken at each of the four quadrants of a petri dish by rotating the sample at angles of 90°. The spectrometer was recalibrated to remove the baseline at every 10 samples with white Spectralon (LabSphere, North Sutton, USA). An average reflectance spectral curve was obtained for each sample for further transformation and modeling.

Pre-processing transformations

The laboratory-measured TC content in the five fields was positively skewed (Table II). In order to reduce the skewness of the TC distributions and model standard error of prediction (Bellon-Maurel *et al.*, 2010), the TC data were logarithm-transformed to approximate normal distributions. Before constructing models, two pre-processing transformations were applied to the soil reflectance curves. For all VNIR spectra, we used the pre-processing methods that worked best in a previous study conducted in Florida (Vasques *et al.*, 2008). Firstly, the reflectance curves were smoothed across a moving window of 9 nm using the Savitzky-Golay algorithm with a third-order polynomial to reduce the random noise (Savitzky and Golay, 1964). Secondly, the first-degree Savitzky-Golay derivative, along with a search window of seven measurements and second-order polynomial, was applied to the smoothed curves.

Regression methods

In order to compare the predictive performance of the three different VNIR diffuse reflectance models, the whole soil TC dataset ($n = 560$) was randomly split in-

TABLE II

Descriptive statistics[a)] of measured soil total carbon (TC) (original values of TC and logarithm-transformed values of TC (logTC)) of the five study fields

Dataset	n	TC						logTC			
		Min	Median	Mean	Max	CV	Skew	Mean	SD	CV	Skew
		%									
Whole	560	0.31	1.04	1.18	3.55	0.55	0.99	0.01	0.24	30.56	0.06
Field 1	112	0.32	0.56	0.59	1.12	0.28	1.02	−0.24	0.11	−0.47	0.37
Field 2	112	0.70	1.63	1.77	3.35	0.36	0.78	0.22	0.15	0.68	0.15
Field 3	112	0.31	0.62	0.68	2.32	0.42	3.28	−0.20	0.14	−0.70	1.28
Field 4	112	0.56	1.05	1.10	2.84	0.30	2.25	0.03	0.11	4.48	0.74
Field 5	112	1.02	1.69	1.76	3.55	0.26	0.80	0.23	0.11	0.47	0.09
CAL	392	0.33	1.02	1.17	3.55	0.55	1.06	0.01	0.23	31.77	0.11
VAL	168	0.31	1.07	1.19	3.21	0.55	0.86	0.01	0.24	28.53	−0.02
SUB-W	112	0.32	1.04	1.20	2.85	0.53	0.69	0.02	0.24	15.29	−0.04

[a)]CAL = the dataset used to calibrate the models; VAL = the dataset used to validate the models; SUB-W = the 112 observations randomly chosen from the five fields (Fig. 1); n = number of observations; Min = minimum; Max = maximum; SD = standard deviation; CV = coefficient of variation; Skew = skewness.

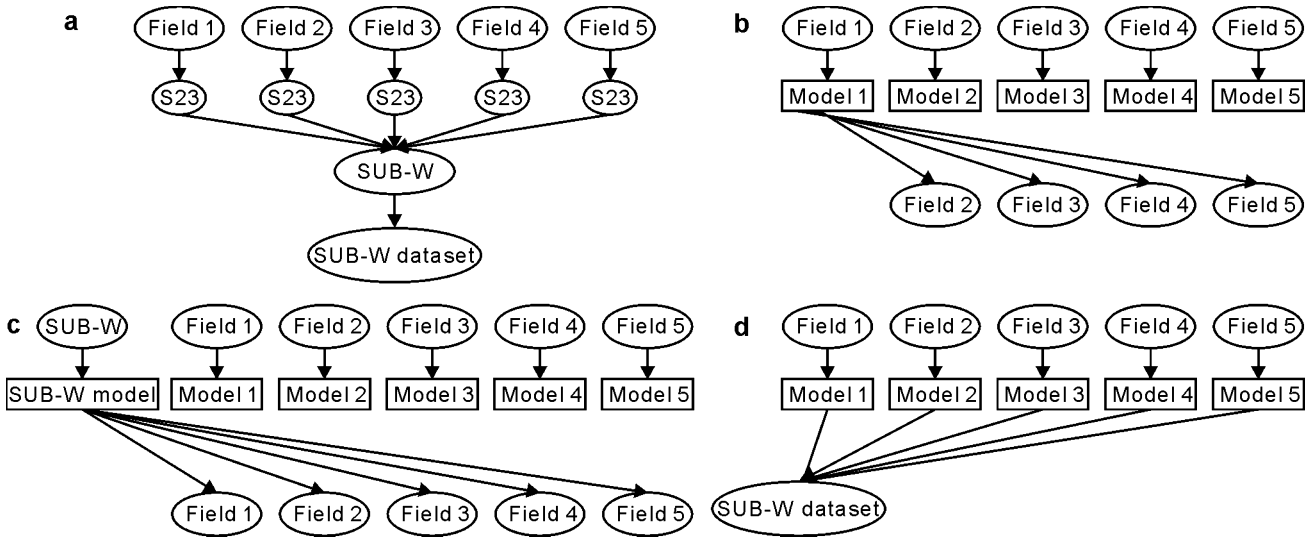

Fig. 1 Principle scheme of the transferability and scalability analysis: the sample source of a pooled subset-whole (SUB-W) dataset (the 112 observations randomly chosen from the five study fields) (a); transferability at field scale (b); downscaling analysis (c); upscaling analysis (d). S23 = the 23 samples randomly chosen from each of the five fields to calibrate the regional model.

to a calibration (CAL) dataset (70%) with $n = 392$ and a validation (VAL) dataset (30%) with $n = 168$. Three different multivariate regression methods were applied to develop the spectral models that were consequently evaluated to predict soil TC using the VNIR spectra: partial least squares regression (PLSR) (Martens and Næs, 1989), support vector machine (SVM) (Vapnik, 2000), and random forest (RF) (Breiman, 2001). The PLSR approach is well suited for the prediction of regression models with a large number of highly collinear predictor variables (Garthwaite, 1994). In this method, the target (*e.g.*, TC) and predictor variables (*e.g.*, spectral data) are simultaneously decomposed into orthogonal principal components (PCs) and a set of specific loadings such that these PCs explain the maximum possible covariance between the target and predictor variables. The predictions were obtained by multiple linear regression of the target variable on the selected PCs, and the number of PCs was chosen by minimizing the RMSE of cross-validation on the CAL dataset. We employed 15 PCs, which represented over 90% of the variation of soil samples. The SVM method was originally proposed by Vapnik (2000) within the area of the statistical learning theory. To address nonlinearity in the input-output data relationships, the SVM method employs a kernel to project the data into a high-dimensional space before performing the regression. In this study, the radial basis function kernel was applied, and a grid search was performed to find the best choices for the parameters for the kernel. The RF method is a learning ensemble consisting of the bagging of unpruned decision tree learners, with randomized selection of predictor variables at each split (Breiman, 2001). All three regression methods were employed to relate the spectral data to the logTC data. Firstly, the leave-one-out (LOO) cross-validation was employed to evaluate the model performance of the CAL dataset. Secondly, an independent validation was used to assess the model performance using the VAL dataset.

The coefficient of determination (R^2) was used as the goodness-of-fit statistic. The RMSE, residual prediction deviation (RPD) (Williams, 1987), ratio of prediction error to interquartile distance (RPIQ) (Bellon-Maurel *et al.*, 2010), and bias (Davies and Fearn, 2006) were provided as complementary error statistics to evaluate the performance of different prediction models.

Model transferability and scalability analysis

In this study, only PLSR and SVM regression models were used in the model transferability and scaling analysis to exemplify the effects of a method that models linear relationships (PLSR) and one that models nonlinear relationships (SVM). Since the behaviors of SVM and RF were very similar in terms of transferability and scalability, only the results derived from SVM were used.

In this study, we adopted the definitions of model transfer, scale transformation, and up-/downscaling as provided by Turner *et al.* (1989), Wu *et al.* (2006), and Blöschl and Sivapalan (1995), respectively. Hence, 'transferability' denotes the transfer (or application) of a VNIR-based soil TC prediction model (Models 1 to 5) developed (calibrated) and validated at one field site (Fields 1, 2, 3, 4, and 5, respectively) to another field site (Fields 1, 2, 3, 4, and 5, respectively) (Fig. 1b). Note that all field sites were of equal

size and contained the same number of observations to build models. This process entails the scaling of neither grain size nor extent and is commonly denoted as 'extrapolation' (Wu *et al.*, 2006). The five models developed for the five fields were denoted as Models 1 to 5. The model performance at calibration sites was assessed using the LOO cross-validation reporting R^2 and RMSE, and the transferability of the models was assessed using R^2, RMSE, RPD, and RPIQ.

In this paper, 'scalability' denotes a change in the extent (size) of the geographic area represented by the models, upscaling refers to an escalation of the area (*i.e.*, from smaller to larger extent), and downscaling refers to the contraction of the area (*i.e.*, from larger to smaller extent) after Wu *et al.* (2006) (Fig. 1c, d). To assess the downscaling behavior, the regional subset-whole (SUB-W) models representing the pooled field areal coverage of approximately 1.25 km^2 were applied to each of the five fields (each approximately 0.25 km^2 in area) (Fig. 1c). On the contrary, to assess the upscaling performance, the models developed using PLSR and SVM for each of the five fields were applied to the regional SUB-W dataset (Fig. 1d). To exclude the impact of the observation size in the scaling procedure, it was kept constant at $n = 112$ for the SUB-W model and for each of the five field models during the scaling procedure. The same error statistics as outlined above were used to evaluate the scaling behavior of the TC models.

In the scaling analysis, a SUB-W dataset was created (with $n = 112$) from the five field datasets (with $n = 560$). The observation size of the SUB-W dataset was equal to that of each field, eliminating any bias or negative effects on the comparative analysis. The SUB-W dataset was randomly selected from the pooled ($112 \times 5 = 560$) field dataset with 1/5 random samples from each of the five fields. The selection process of the SUB-W dataset is shown in Fig. 1a. The models calibrated by the spectral SUB-W data were evaluated using the LOO cross-validation.

Similarity of soil environmental conditions among fields and across scales

To examine the constraining effect of soil environmental conditions on the transfer of the soil spectral models across fields and scales, the similarities among fields and across scales were characterized by the SCORPAN factors (McBratney *et al.*, 2003). The SCORPAN factors include soil characteristics (S), climate (C), organism, vegetation, and land use (O), relief (R), parent material (P), age (A), and space (N). Assuming the homology of the soil-forming factors between a calibration area and the region of interest, Mallavan *et al.* (2010) proposed that the smaller the taxonomic distances of the SCORPAN factors, the more similar the soil characteristics. The Gower similarity coefficient (Gower, 1971; Booth *et al.*, 1987), as outlined by Mallavan *et al.* (2010), was employed to measure the similarity in soil-forming factors among fields:

$$S_{ij} = \frac{1}{p}\sum_{k=1}^{p}\left(1 - \frac{|x_{ik} - x_{jk}|}{\text{range } k}\right) \quad (1)$$

where S_{ij} is the Gower similarity coefficient between sites i and j; k represents the SCORPAN variables; p is the number of variables; range k is the value range of variable k in the whole study area. Important variables that were included in the similarity analysis are shown in Table I. The Gower similarity coefficients were compared with the model prediction performance.

RESULTS AND DISCUSSION

Descriptive statistics

The TC content of the total 560 soil samples, collected from all five fields, showed a positively skewed distribution, with a mean of 1.18%, a median of 1.04%, and a range between 0.31% and 3.55% (Table II). The minimum and maximum values of the logarithm-transformed TC were −0.5% and 0.55%, respectively, with a mean of 0.01% and median of 0.02%. The descriptive statistics of the CAL, VAL, and SUB-W datasets were similar to the whole dataset ($n = 560$), indicating that all of them appropriately represented the population of the whole study area. Among all soil samples, the highest TC values occurred in Fields 2 and 5.

The soil TC mean and median (%) were the highest in Fields 2 and 5 and declined in Fields 4, 3, and 1 (Table II), resembling a hydrologic gradient as indicated by the available water-holding capacity (AWHC), which was the highest in Fields 4 and 2 and declined in Fields 5, 3, and 1 (Table I). The trajectory of the soil TC values also mirrors the land use/land cover gradient with the highest soil TC found under the mesic upland forest and rangeland and the lowest soil TC occurring under the xeric upland forest.

Assessment of prediction performance for soil TC

The results of the soil logTC prediction of the models derived from CAL and VAL using the three different regression methods are shown in Table III. The performance of the SVM and RF models was very similar in both LOO cross-validation and validation modes.

In the LOO cross-validation mode, the PLSR models performed slightly better, with the highest R^2 (0.88) and the lowest RMSE (0.08%) values, as compared with the SVM (with $R^2 = 0.87$ and RMSE = 0.09%) and RF (with $R^2 = 0.87$ and RMSE = 0.08%) models. In the validation mode, the PLSR models had a smaller bias, although the R^2 value was slightly lower, compared with the other two methods. The predicted *vs.* observed TC in the VAL dataset matched well for all three methods with the values aligned close to the 1:1 line (Fig. 2). However, the SVM and RF models tended to slightly overpredict when logTC was smaller than zero and underpredict in the high logTC data range. The data of the PLSR models are almost homogenously distributed along the 1:1 line. The high RPD (≥ 2.70) and RPIQ (≥ 4.54) values for all three models confirmed their excellent prediction accuracies.

Transferability and scalability analysis

The soil TC prediction performances declined at varying degrees when the PLSR and SVM models were transferred to the field outside the calibration geographical area (Fig. 1b), downscaled (Fig. 1c), and upscaled (Fig. 1d). Overall, the PLSR model showed better downscaling and upscaling performance than the SVM model. For the SVM model, only downscaling produced acceptable TC predictions. Although the SUB-W model (with $n = 112$) prediction performance (R^2 of 0.82 for PLSR and 0.84 for SVM and RMSE of 0.10% for both PLSR and SVM) decreased slightly (Table IV) compared to that of the calibration models (with $n = 392$) (Table III), the SUB-W model could still effectively represent the VNIR-TC relationship at the regional scale. The goodness-of-fit statistic R^2 for Models 1–5 at the field scale ranged from 0.46 to 0.69 (PLSR) and 0.33 to 0.59 (SVM) (Table IV), indicating that the performance of the models developed at the field scale was not as good as those developed at the regional scale, using the same size of calibration sample ($n = 112$).

Transferability of soil TC prediction models among field sites. In order to test the model transferability at field scale, the PLSR and SVM models developed and cross-validated at one field were applied to the ot-

TABLE III

Summary statistics[a)] for the spectral models of logarithm-transformed soil total carbon (logTC) produced using the partial least squares regression (PLSR), support vector machine (SVM), and random forest (RF) methods derived from calibration dataset (CAL) using 70% of all samples ($n = 392$) and validation dataset (VAL) using 30% of all samples ($n = 168$)

Regression method	Leave-one-out cross-validation using CAL		Validation using VAL				
	R^2	RMSE	R^2	Bias	RMSE	RPD	RPIQ
		%		%			
PLSR	0.88	0.08	0.86	0.004	0.09	2.70	4.54
SVM	0.87	0.09	0.88	0.01	0.09	2.78	4.67
RF	0.87	0.08	0.88	0.01	0.09	2.80	4.70

[a)] R^2 = coefficient of determination; RMSE = root mean squared error; RPD = residual prediction deviation; RPIQ = ratio of prediction error to interquartile range.

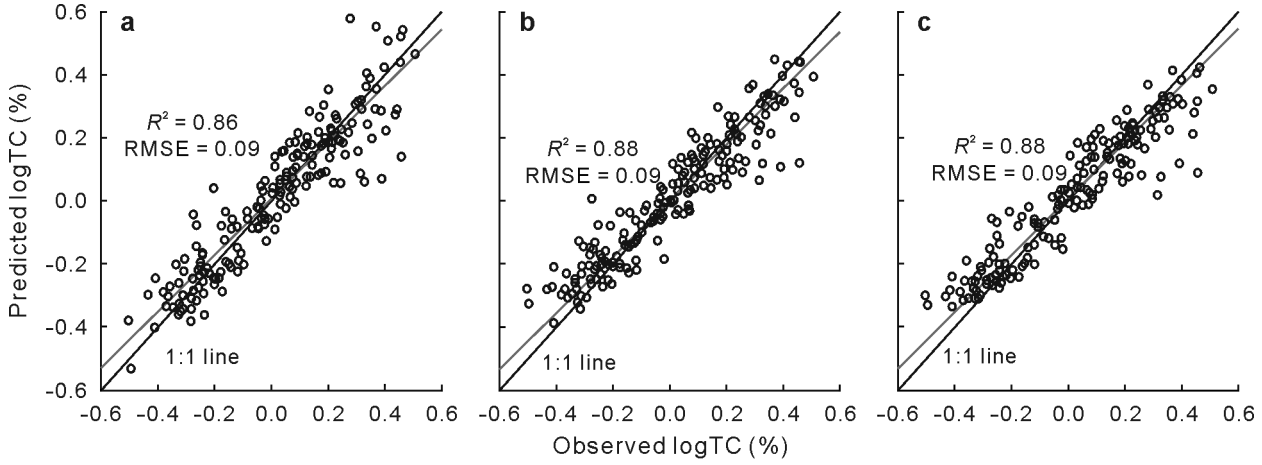

Fig. 2 Predicted soil total carbon (TC) *vs.* observed logarithm-transformed soil TC (logTC) of validation dataset derived from three methods: partial least square regression (PLSR) (a), support vector machine (SVM) (b), and random forest (RF) (c). R^2 = coefficient of determination; RMSE = root mean squared error.

TABLE IV

Summary statistics[a)] of the leave-one-out cross-validation for the partial least squares regression (PLSR) and support vector machine (SVM) models of logarithm-transformed soil total carbon (logTC) developed in five study field datasets and the subset-whole (SUB-W) dataset (the 112 observations randomly chosen from the five study fields)

Model	PLSR		SVM	
	R^2	RMSE	R^2	RMSE
		%		%
SUB-W model	0.82	0.10	0.84	0.10
Field dataset				
Model 1	0.69	0.06	0.55	0.08
Model 2	0.62	0.10	0.59	0.11
Model 3	0.46	0.10	0.33	0.11
Model 4	0.56	0.07	0.59	0.08
Model 5	0.61	0.07	0.52	0.08

[a)] R^2 = coefficient of determination; RMSE = root mean squared deviations.

her four fields. The results of the transferability of PLSR and SVM models are summarized in Tables V and VI, respectively. The R^2 values of the PLSR models were below 0.39, except when Model 1 was applied to predict the soil TC of Field 2 ($R^2 = 0.53$) and Field 3 ($R^2 = 0.51$). The TC contents of Field 1 predicted by Models 3, 4, and 5 were severely biased (0.37%). The high bias and RMSE values with low RPD and RPIQ values suggest that the transferability among field sites was severely constrained using the PLSR models. The R^2 values derived from the SVM models were significantly lower than those derived from the PLSR models, with several values even smaller than 0.01, indicating that the SVM model transferability among field sites was severely constrained (Table VI). The limited transferability using the SVM model was confirmed by the high bias and prediction error (Table VI). Interestingly, when the two models calibrated in different fields were transferred to each other, they behaved in different ways. For example, although Model 1 predicted the TC content in Field 2 well with R^2 of 0.53, the performance of Model 2 was poor when transferred to Field 1.

Downscaling of soil TC prediction models. To test the downscaling performance, the regional pooled model (SUB-W) was applied to predict the soil TC of the five fields (Table VII). Both PLSR and SVM models successfully predicted the TC of Fields 2 and 4 with R^2 larger than 0.47, but the performance in Fields 1 and 3 was limited with R^2 smaller than 0.42. The PLSR and SVM models predicted the TC of Field 5 differently. The PLSR model fit was constrained (R^2

TABLE V

Transferability[a)] of partial least squares regression (PLSR) models developed in one of the five study fields to predict the logarithm-transformed soil total carbon (logTC) of the other four fields

Model	Validation dataset ($n = 112$)	R^2	Bias	RMSE	RPD	RPIQ
			%			
Model 1	Field 2	0.53	−0.15	0.19	0.83	1.27
	Field 3	0.51	−0.16	0.21	0.66	0.68
	Field 4	0.17	−0.34	0.36	0.32	0.33
	Field 5	0.11	0.01	0.17	0.64	0.95
Model 2	Field 1	0.15	0.01	0.11	0.99	1.36
	Field 3	0.39	−0.10	0.18	0.74	0.77
	Field 4	0.15	−0.01	0.17	0.67	0.70
	Field 5	0.17	−0.23	0.27	0.40	0.59
Model 3	Field 1	0.12	0.28	0.31	0.37	0.51
	Field 2	0.09	−0.46	0.58	0.28	0.42
	Field 4	0.02	−0.21	0.34	0.34	0.35
	Field 5	0.02	0.01	0.20	0.54	0.80
Model 4	Field 1	0.34	0.34	0.35	0.32	0.44
	Field 2	0.29	0.05	0.15	1.09	1.67
	Field 3	0.32	0.19	0.23	0.59	0.61
	Field 5	0.34	0.19	0.21	0.51	0.75
Model 5	Field 1	0.24	0.37	0.39	0.29	0.41
	Field 2	0.28	0.05	0.16	0.98	1.49
	Field 3	0.25	−0.23	0.28	0.48	0.50
	Field 4	0.22	0.07	0.14	0.82	0.85

[a)] R^2 = coefficient of determination; RMSE = root mean squared error; RPD = residual prediction deviation; RPIQ = ratio of prediction error to interquartile range.

TABLE VI

Transferability[a)] of support vector machine (SVM) models predicting logarithm-transformed soil total carbon (logTC) developed in one of the five study fields to predict the soil logTC of the other four fields

Model	Validation dataset ($n = 112$)	R^2	Bias	RMSE	RPD	RPIQ
			%			
Model 1	Field 2	0.13	−0.47	0.49	0.32	0.49
	Field 3	0.12	−0.06	0.15	0.93	0.96
	Field 4	< 0.01	−0.28	0.30	0.38	0.39
	Field 5	0.31	−0.48	0.49	0.22	0.33
Model 2	Field 1	0.06	0.39	0.40	0.28	0.39
	Field 3	0.17	0.37	0.39	0.35	0.36
	Field 4	< 0.01	0.14	0.18	0.62	0.65
	Field 5	0.21	−0.06	0.12	0.88	1.30
Model 3	Field 1	0.33	0.09	0.14	0.81	1.12
	Field 2	0.01	−0.38	0.41	0.39	0.60
	Field 4	0.02	−0.19	0.22	0.52	0.54
	Field 5	0.27	−0.39	0.41	0.27	0.40
Model 4	Field 1	< 0.01	0.28	0.30	0.38	0.52
	Field 2	< 0.01	−0.18	0.24	0.67	1.02
	Field 3	0.06	0.23	0.27	0.51	0.53
	Field 5	0.18	−0.20	0.22	0.49	0.72
Model 5	Field 1	0.04	0.46	0.47	0.24	0.33
	Field 2	< 0.01	0.00	0.16	1.00	1.53
	Field 3	0.05	0.41	0.43	0.32	0.33
	Field 4	< 0.00	0.19	0.22	0.51	0.53

[a)] R^2 = coefficient of determination; RMSE = root mean squared error; RPD = residual prediction deviation; RPIQ = ratio of prediction error to interquartile range.

TABLE VII

Downscaling performance statistics[a)] of the soil logarithm-transformed soil total carbon (logTC)-predicting partial least squares regression (PLSR) and support vector machine (SVM) models developed at the regional subset-whole (SUB-W) scale (the 112 observations randomly chosen from the five study fields) to predict samples of the five field

SUB-W model	Validation dataset ($n = 112$)	R^2	Bias	RMSE	RPD	RPIQ
			%			
PLSR	Field 1	0.42	< 0.01	0.11	1.07	1.47
	Field 2	0.47	−0.02	0.13	1.27	1.93
	Field 3	0.32	0.07	0.16	0.87	0.90
	Field 4	0.51	−0.04	0.10	1.10	1.14
	Field 5	0.20	−0.03	0.16	0.68	1.00
SVM	Field 1	0.35	0.08	0.12	0.92	1.26
	Field 2	0.55	−0.08	0.13	1.19	1.81
	Field 3	0.26	0.08	0.14	0.95	0.99
	Field 4	0.65	0.01	0.07	1.63	1.69
	Field 5	0.51	−0.03	0.08	1.32	1.94

[a)] R^2 = coefficient of determination; RMSE = root mean squared error; RPD = residual prediction deviation; RPIQ = ratio of prediction error to interquartile range.

(= 0.20), while the SVM model performed well (R^2 = 0.50). In general, the RPD and RPIQ values of both PLSR and SVM models suggest that their downscaling behaviors were acceptable, but far from stellar performance of models. It is important to note that the downscaling behavior from the regional to field scale differed substantially among fields.

Upscaling of soil carbon prediction models. Table VIII shows the upscaling behavior of the five field models (Models 1 to 5) to predict the TC of SUB-W at the regional scale. The performances of the five PLSR field-scale models, except for Model 3, were as good as the downscaled models using PLSR. The R^2 of the upscaled PLSR Models 1, 2, and 4 were all above 0.50, and the RPIQ values higher than 1.84, suggesting a moderately good upscaling performance (Table VIII). The PLSR Model 3 failed to predict the TC of SUB-W, which was due to its poor performance in the LOO cross-validation mode (Table IV). Although the PLSR Models 1, 2, 4, and 5 had high R^2 values, the large bias and RMSE values showed that upscaling did not perform as well as downscaling. In contrast, the upscaling of SVM models was very poor. The field-scale SVM models predicted the TC of the SUB-W dataset with R^2 below 0.19 and RMSE higher than 0.23%, and RPIQ values slightly lower than those predicted by the PLSR models.

Gower similarity coefficient

The Gower imilarity coefficients were all above 0.50 (Table IX). The coefficients between the SUB-W dataset and each of the five fields were higher than those between the five fields. Correspondingly, the downscaling performance of the SUB-W models using the PLSR and SVM methods outperformed the field models when transferred to the other fields.

TABLE VIII

Upscaling performance[a)] of the logarithm-transformed soil total carbon (logTC)-predicting partial least squares regression (PLSR) and support vector machine (SVM) models developed at the field scale to predict samples at the regional subset-whole (SUB-W) scale (the 112 observations randomly chosen from the five study fields)

Model	Validation dataset ($n = 112$)	R^2	Bias	RMSE	RPD	RPIQ
PLSR			%			
Model 1	SUB-W	0.53	−0.12	0.22	1.09	1.97
Model 2	SUB-W	0.64	−0.03	0.15	1.58	2.87
Model 3	SUB-W	< 0.01	−0.17	0.50	0.48	0.86
Model 4	SUB-W	0.57	0.18	0.23	1.02	1.84
Model 5	SUB-W	0.36	0.06	0.23	1.02	1.86
SVM						
Model 1	SUB-W	0.08	−0.27	0.35	0.67	1.22
Model 2	SUB-W	0.19	0.15	0.27	0.90	1.63
Model 3	SUB-W	0.04	−0.18	0.29	0.81	1.48
Model 4	SUB-W	0.11	0.03	0.23	1.05	1.91
Model 5	SUB-W	0.15	0.21	0.31	0.77	1.40

[a)] R^2 = coefficient of determination; RMSE = root mean squared error; RPD = residual prediction deviation; RPIQ = ratio of prediction error to interquartile range.

TABLE IX

Gower similarity coefficients of soil environmental factors among the five study fields and across scales[a)]

Dataset	Field 1	Field 2	Field 3	Field 4	Field 5	SUB-W
Field 1	1.00	0.64	0.68	0.73	0.71	0.78
Field 2	–	1.00	0.72	0.76	0.66	0.81
Field 3	–	–	1.00	0.62	0.69	0.80
Field 4	–	–	–	1.00	0.63	0.81
Field 5	–	–	–	–	1.00	0.80
SUB-W	–	–	–	–	–	1.00

[a)] The regional subset-whole (SUB-W) dataset is the 112 observations randomly chosen from the five study fields.

Prediction performance of spectral prediction models

The soil TC predictions derived from all three multivariate statistical methods (PLSR, SVM, and RF) for the five study field showed good performance within the observation range of 0.31% to 3.55%, which was slightly narrower than in other studies (McCarty *et al.*, 2002; Vasques *et al.*, 2008; Sarkhot *et al.*, 2011). Brown *et al.* (2005) found that the VNIR models developed using the boosted regression trees (BRT) outperformed PLSR to predict SOC and soil TC, while McDowell *et al.* (2012b) found no significant difference among the

PLSR and RF ensemble regression trees to predict TC of Hawaiian soils, USA. The global SOC models from the VNIR data were derived with an R^2 of 0.89 and an RPD of 3.3 in the validation mode (Viscarra Rossel *et al.*, 2016).

Minasny and McBratney (2008) and Minasny *et al.* (2009) found excellent predictions for SOC and soil TC in Australia, using the regression rules (Cubist) approach. In contrast, Vasques *et al.* (2010) identified the SOC predictions made using the ensemble regression trees as more accurate than those derived from PLSR in an investigation in Florida. This points to the fact that depending on the geographic soil region, one method might outperform several others to predict SOC or soil TC from the VNIR spectra.

Factors that impact the transferability and scalability of prediction models

Overall, the PLSR models showed better transferability and scalability than the SVM models. This implies that the linear relationships between the VNIR spectra and soil TC (quantified by PLSR) were more pronounced than the nonlinear, complex relationships (quantified by SVM). The reason that the transferability and scalability of soil prediction models were constrained might be explained by the differences in: i) the spectral data domain space, ii) the soil attribute domain space, iii) the methods that determine the internal model structure of VNIR-TC relationships, and iv) the environmental domain space of the attributes that control soil carbon dynamics (*i.e.*, SCORPAN factors). In this study, the number of observations to build models was kept constant in the experimental setup to avoid bias, and comparison between models was allowed from a statistical perspective. One factor that might have impacted the up- and downscaling behaviors of TC models was the sampling density that differed among the field sites (each approximately 0.25 km^2 with a density of observations of approximately 448 per km^2) and the SUB-W regional dataset (representing an area of approximately 1 km^2 with a density of observations of approximately 112 per km^2). In essence, the sample numbers of the field and regional models were the same ($n = 112$), while the geographical area of the regional model was five times larger than that of all field models. According to McBratney (1998), upscaling to larger geographic extent (*i.e.*, increase in study area) inherently increases the variance and upper and lower bounds of both soil properties and soil-forming factors. This effect of increasing variances was also observed by Vasques *et al.*(2012b), who upscaled soil TC from fine scale (5.6 km^2) to intermediate scale (3 500 km^2), and then to coarse scale (150 000 km^2) in Florida. The soil scaling studies using the VNIR data have been rarely presented in the literature. One of the few studies to downscale the SOC-VNIR models, *i.e.*, to predict SOC at local scale using the VNIR spectra and a large spectral library was presented by Guerrero *et al.* (2016). However, the well-performing local models for SOC ($R^2 > 0.95$ and RPIQ > 5.48) was only achieved through spiking local samples and an extra-weighting procedure. Guerrero *et al.* (2016) suggested that the large spectral libraries are not needed for the local assessment of SOC because they favor the development of small-sized spectral libraries customized toward a local scale.

Spectral data domain. The transferability and scalability of models might also be dependent on the spectral data domain. The VNIR models to predict TC selected variables in the spectral regions of the absorption features of C–H, N–H, and O–H groups, similar to the VNIR models presented by Vasques *et al.* (2008, 2009, 2010). These spectral signatures are produced by the overtones and combinations of absorption molecular vibrations (*e.g.*, C–H, O–H, H_2O, and CO_3^-) in the MIR regions (Brown *et al.*, 2005). The features associated with TC can be masked or distorted by the iron oxides and secondary clays, which are commonly found in soils (Hunt, 1989; Clark, 1999). This alludes to a critical issue of VNIR modeling that other soil properties, such as texture, nutrient content, and minerals, might mask or interfere with the prediction of a given property of interest (*e.g.*, soil TC); thus, the transferability of models is affected. In this study, the soil texture differed only slightly among the five sites with sand in the range between 90.8% to 98.6% and clay between 1.2% to 5.2%. Hence, the effect of soil texture imposed on the TC spectral signatures was likely minor. Since soil samples were dried and scanned under controlled laboratory conditions, the impact of differences in soil moisture among sites were excluded from this study. The soil suborders, including Entisol (Psamment), Ultisol (Aquult and Udult), Inceptisol (Udept), and Spodosol (Aquod), differed among sites (Table I), suggesting that the mineralogy, sesquioxides, and other chemical and physical soil properties differed substantially among sites. This might have constrained the transferability and scalability of the VNIR-based TC prediction models because of the masking or distortion effects of the spectral data domain.

Soil attribute domain. The soil attribute domain, *i.e.*, the upper and lower bounds and dispersion of soil TC used to build the spectral prediction models (Table II), might explain some of the transferability

and scalability behavior of models. Typically, the soil attribute domain space expands as the geographic size of the modeled region increases (Grunwald *et al.*, 2011). The range of the soil TC values of CAL, VAL, and SUB-W datasets matched reasonably well a minimum of 0.31% and a maximum of 3.55% of the whole data set. However, the differences in soil TC among field sites were profound (Table II). The transferability of a TC prediction model to other sites might lead to an extrapolation outside the soil attribute observation range of the original field data, thereby affecting its performance. Ideally, the boundary conditions of attributes used for model development of a transfer function (or calibration spectral model) matches the boundary conditions of a transfer dataset. Brown *et al.* (2005) demonstrated the implications of spectral model transfer to predict soil carbon in other fields in Montana, USA, where the SOC values differed widely among field sites (a minimum of 1.93 g kg^{-1} and a maximum of 15.82 g kg^{-1}). They found that PLSR could effectively model individual field sites located within the same physiographic region. However, when they attempted to predict SOC for each of the six sites using the remaining five sites for calibration, the models failed completely at two of the six sites and gave inconsistent results at a third site, despite prescreening for spectral similarity.

In this study, Models 1, 3, and 4, which resembled the TC range of SUB-W most closely with the minimum TC of 0.32% and the maximum TC of 2.85%, did not show persistent responses in terms of transferability based on PLSR (Table V). For example, Model 3 (Psamment on pineland) failed to transfer well to Field 4, whereas the opposite was found for the transfer behavior of Model 4 (Udult on improved pasture) to Field 3. The models that exceeded the upper bound TC of SUB-W (Model 2, diverse soil suborders on mesic upland forest and Model 5, Aquod on rangeland) transferred somewhat equally well to other fields, suggesting that the attribute domain range did not substantially impact model transferability. These findings were confounded in downscaling mode. The SUB-W Model with the minimum TC of 0.32% and the maximum TC of 2.85%) substantially degraded more when applied to Field 5 with a mismatched attribute domain space (with the minimum TC of 1.02% and maximum TC of 3.55%) than when applied to the field sites with similar attribute domain spaces (*e.g.*, Field 4 with the minimum TC of 0.56% and the maximum TC of 2.84%) or to the field sites with narrower attribute domain spaces (*e.g.*, Field 1 with the minimum TC of 0.32% and maximum TC of 1.12%) (Tables II and VII). In upscaling mode, the models that showed a wider or matching attribute domain space (Models 1, 2, and 4) compared to SUB-W performed reasonably well to scale, whereas Model 3 that matched the TC upper and lower bounds of SUB-W failed to scale, and Model 5 degraded somewhat when upscaled to SUB-W (Tables II and VIII). These findings substantiate that no clear conclusions emerge on linkage of the attribute domain boundaries to the transferability and scalability of models.

Besides the upper and lower bounds of attributes that matter for successful model transfer and scaling, it is the internal variability (variance) of soil attributes that potentially impacts the behaviors. Addiscott *et al.* (1995) pointed out that an increase in parameter variance might cause problems by interacting with the linearity/nonlinearity in the process represented by the model. McBratney (1998) and Grunwald *et al.* (2011) asserted that an increase in the variance of soil attributes could affect the model building process, transferability, and scalability of soil properties. In this study, the coefficient of variation (CV) ranged from 0.26% (Field 5) to 0.42% (Field 3), which was lower than in the pooled sets (0.53% and 0.55% in the SUB-W and whole datasets, respectively). The lower CV in Field 5 somewhat limited the transferability of TC models to other field sites more than the transferability of other models to field sites (Table V). It is interesting to note that Model 3 (Psamment on pineland), which had the highest variability in TC among field sites, performed worst in terms of transferability to other field sites. In contrast, the SUB-W model, which had higher variability in TC than Model 3, performed substantially better, suggesting that the internal variability in attributes are one, but perhaps not the most important controlling factor that limits model transferability and scalability. The downscaling performance of the SUV-W model with the highest CV of 0.53% had relatively less impact on Field 3 (with an intermediate CV of 0.42%), but substantially degraded when applied to Field 5 (with the lowest CV of 0.26%) for the PLSR and SVM models (Table VII). These findings suggest that although the SUV-W model represented the variability in the TC of the five fields, it severely degraded the downscaling of those fields that were more homogeneous in soil TC (*e.g.*, Field 5). The upscaling to the SUB-W model failed for Model 3 and was severely muted for Model 5, whereas Models 1, 2, and 4 were relatively less impacted. This confirms the assertion that the variance in TC plays a role in the scalability of models; however, other substantial factors confound these findings.

The spiking of a spectral model with local samples has been suggested to improve soil predictions (Sankey *et al.*, 2008; Wetterlind and Stenberg, 2010). However, its success is highly dependent on multiple factors, including the ratio between the number of spike samples and the number of samples in the spectral library, characteristics of the soil attribute and spectral domain spaces, and the methods used to develop soil spectral prediction models. Although the same constraints, mechanisms, and effects influence the spiking and scaling of chemometric models, their aims are inherently different. The spiking process aims to stabilize/improve the predictions of soil properties by adding more observations to the dataset, whereas the scaling process aims to understand the factors and processes affecting the scaling behavior. An example of spiking a local subset to improve the model predictions of SOC was provided by Guerrero *et al.* (2014), with the best results obtained when the spiking dataset contained local samples evenly distributed in the spectral space.

Methods (model types). The regression methods use different strategies to relate predictors (spectral data in this study) and a response variable (soil TC in this study). The underlying strategies for selecting a predictor are different for PLSR and SVM, as described above, affecting their transfer and scale responses. If the internal model structure that describes the relationship between the spectral predictors and soil TC is not stable when it is scaled, it suggests a scale-variant behavior. As expected, the PLSR (Fig. 3a) and SVM (Fig. 3b) models showed differences in the selection of spectral predictors in Models 1–5 and the SUB-W model. Thissen *et al.* (2004) also found major differences in the selection of spectral predictors that are inherent to the modeling process of PLSR and SVM, specifically in cases where the physicochemical composition of the soil samples differs. In this study, the PLSR method was more robust than the SVM method in transferring models among sites. To identify spectral predictors, the PLSR models (Fig. 3a) mainly focused on three regions, including those around 350, 1 860, and 2 200 nm, which represented the reflection region of organic matter (Galvao and Vitorello, 1998), O–H$^-$, water, C–H, C–N, C–O, N–H (Vasques *et al.*, 2008), calcium carbonate (2 206 and 2 341 nm) (Lagacherie *et al.*, 2008), or M–OH and various C–O groups (Brown *et al.*, 2005). On the other hand, the top 50 important spectral wavelengths of the SVM models (Fig. 3b) were found around 670, 1 400, 1 800, and 2 200 nm. In particular, 670 nm is indicative of iron oxide features (McDowell *et al.*, 2012b) and 1 400–1 900 nm are the absorption regions of O–H and water, showing that iron oxide and crystallization of water affected TC predictions. In the VNIR spectral range, only overtones are mapped; this differs from other spectral methods, such as MIR sensing, which respond more directly to the chemical composition of samples (McDowell *et al.*, 2012a, b).

The SVM modeling of SOC using the VNIR data has become popular, as exemplified in some studies conducted in China (Peng *et al.*, 2014), Brazil (Dotto *et al.*, 2017), and Germany (Morellos *et al.*, 2016). Although SVM is advantageous for modeling complex, high-dimensional spectral datasets because it can model nonlinear structures, it performed poorly in terms of model transferability and scalability. This can be explained by the high susceptibility of SVM to overfitting (Hernández *et al.*, 2009). The substantially larger amount of spectral values selected was important in the SVM model compared to the PLSR model suggests overfitting (Fig. 4). Most of the spectral predictors were assigned larger values in the SVM model compared with the PLSR and RF models (Fig. 4). In addition, the nonlinear relationships between the spectral data and TC might be rather weak, limiting its competitiveness when compared to a more robust, linear method such as PLSR (Hernández *et al.*, 2009).

Environmental domain space of attributes. Soil carbon gains/losses have been linked to various environmental factors, such as climate (Hook and Burke, 2000), land use/land cover (John *et al.*, 2005; Rees *et al.*, 2005), soil moisture/hydrology (Vasques *et al.*, 2012b), and topography (Yimer *et al.*, 2006). Mallavan *et al.* (2010) argued that soil attributes correlate consistently with environmental factors, assuming the homology of soil-forming factors. The concept of homosoil asserts that soil TC in an unsampled area can be inferred from the modeled relationships of soil TC and environmental covariates derived from a sampled area, given that both areas are similar in terms of the environmental factors (Minasny and McBratney, 2010).

The environmental factors (*i.e.*, the soil-forming factors) of fields differed widely in terms of topography, climate, parent material, organism/biota, and soils (Table I). The homology among environmental conditions explained the ability to transfer TC models to other field sites and scales in this study (Fig. 5, Table IX). Minasny *et al.* (2009) found that the transfer of the MIR spectral SOC prediction models among three different regions, Queensland, New South Wales, and Victoria, in Australia did not perform well because of

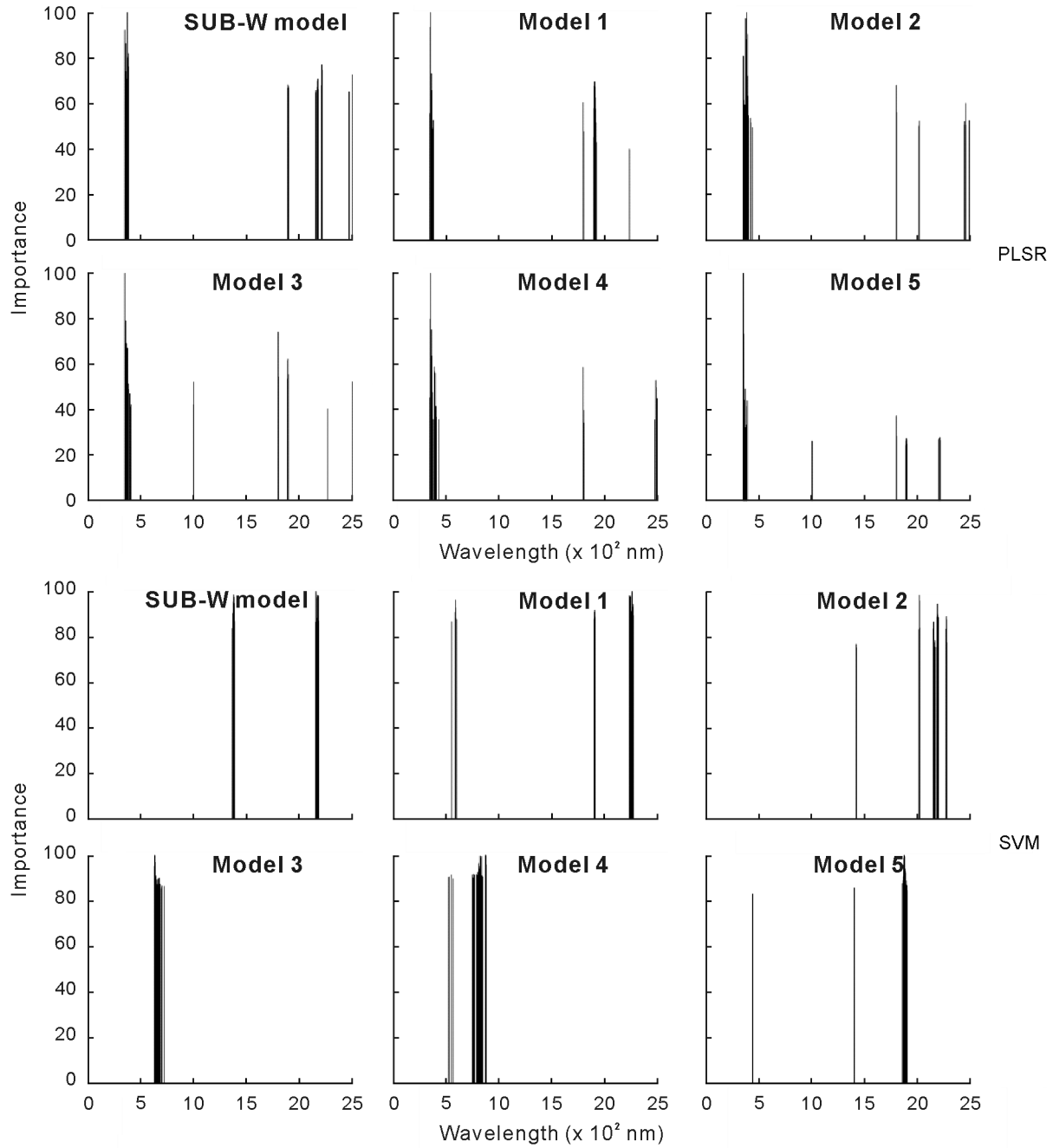

Fig. 3 Importance values of top 50 predictors (wavelengths) derived from the models with the partial least squares regression (PLSR) (a) and support vector machine (SVM) (b) methods. SUB-W model 1 is developed from the pooled regional subset-whole (SUB-W) dataset (112 observations randomly chosen from the five study fields; Models 1, 2, 3, 4, and 5 are developed from the datasets of Fields 1, 2, 3, 4, and 5, respectively.

the differences in the parent material and climate under which the soils have formed. Unfortunately, no explicit similarity analysis of environmental factors was presented in their study, and the relationships between soil environmental factors are not clear (Minasny *et al.*, 2009). Although the R^2 values of the transferred models were still moderate, all models showed a significant bias. Studies investigating both the similarity in soil TC (or other soil properties) among sites and the similarity in environmental factors that form those soil properties are still rare in the soil science literature.

CONCLUSIONS

Although the spectral models used to predict soil TC with three different methods SR, SVM, and RF) were successful in the calibration and validation modes at five different fields nested within a large sand-domi-

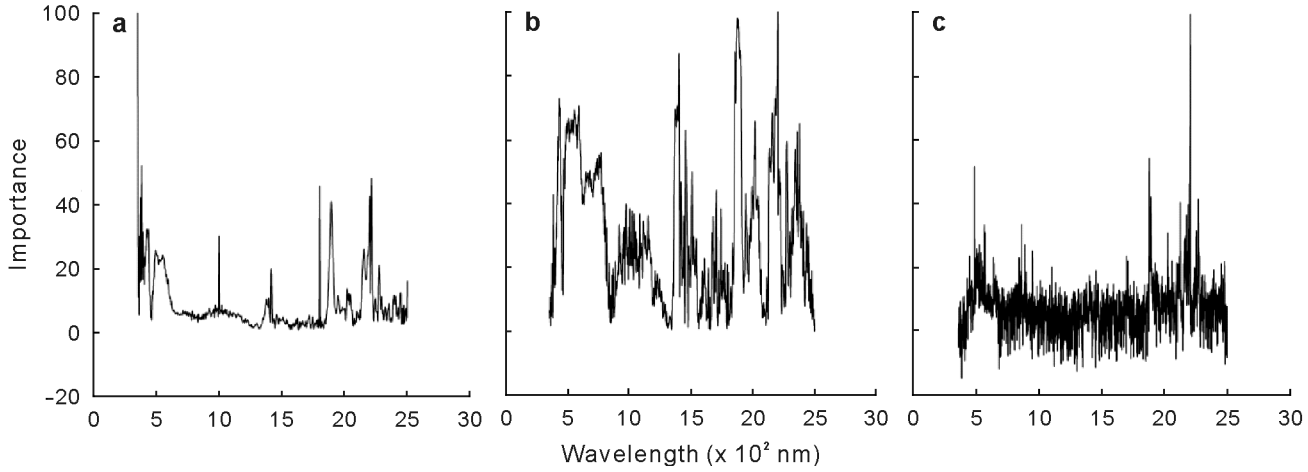

Fig. 4 Variable importance values of the partial least squares regression (PLSR) (a), support vector machine (SVM) (b) and random forest (RF) (c) models derived from the calibration dataset.

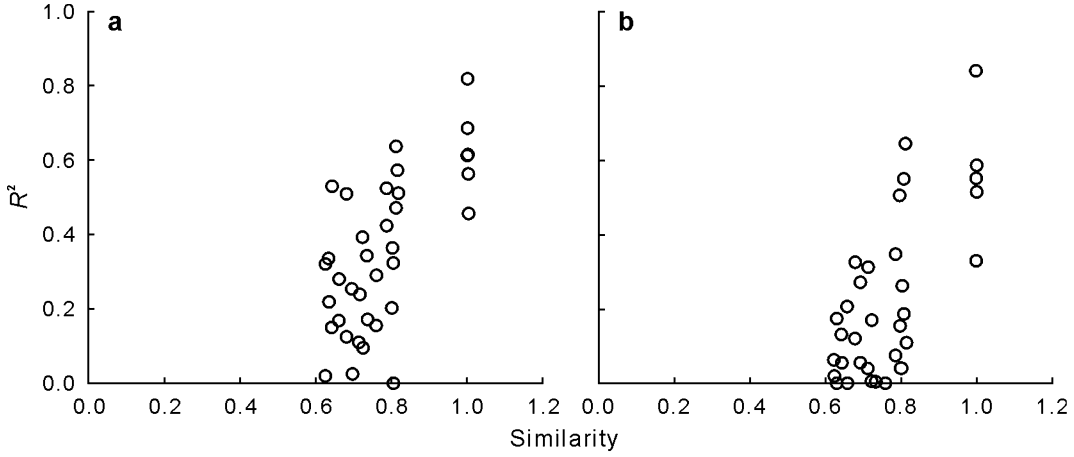

Fig. 5 Coefficient of determination (R^2) of each model transferred to other fields and scales *vs.* the Gower similarity coefficient between the model development field/scale and model application field/scale: partial least squares regression (PLSR) (a) and support vector machine (SVM) (b) models.

nated region in the U.S., the transferability and up- and downscaling of the models were limited by the following factors: i) the spectral data domain space, ii) the soil attribute domain space, iii) the methods that describe VNIR-TC relationships, and iv) environmental domain space of attributes that control soil carbon dynamics. All of these four factors interacted with each other, affecting the transferability of models among fields and the up- and downscaling behaviors of the spectral soil prediction models. Overall, the transferability and scalability of the prediction models derived from PLSR were better and more robust than those derived from SVM. However, no universal trend indicating which of the four limiting factors mentioned above had the maximum constraining impact on the transferability and scalability of models was found. These findings have implications for the development of universal spectral models aiming to predict soil properties for a diverse set of different soils formed in different environmental conditions and covering a wide range of geographic settings throughout the world. Those universal spectral libraries are based on the premise that soil predictions (*e.g.*, soil TC) can be made anywhere because they are built using the soil spectral datasets that exhaustively characterize the attribute feature space. This assertion is limited by the fact that a large number of interacting factors of soils, spectra, and environmental properties are needed to represent the exhaustive sample population, which has not materialized yet. The SUB-W model substantially degraded when applied to local fields in terms of performance. In analogy, a universal spectral library might also suffer from severe degradation for prediction at a local (site-specific) scale. Specifically, if universal soil spectral libraries are created using the crawling approach (*e.g.*, assembling the soil spectral data from the publicly available databases irrespective of quality) or spiking/pooling of random, rather

than strategic, soil spectral data (*e.g.*, based on funded projects that generate data) leading to extremely unbalanced datasets where one geographic region is overrepresented and others underrepresented. Given that several factors can impinge on the empirically derived soil spectral prediction models, as demonstrated by this study, more focus on their applicability and scaling is needed. This study juxtaposed the local and regional predictions, transferability, and scalability of soil TC models derived from the VNIR spectra within a subtropical region in the southeastern U.S. The constraints and limitations of the soil spectral models identified in this study might also be found in other regions as well as spectral libraries that intend to have universal applicability. Furthermore, the stationarity in mean and variance in local (field) calibrations of soil spectral prediction models are usually easier to meet though can have severe effects on the scale-variant behavior of models at escalating spatial scales. The confounding trends in the soil TC up- and downscaling behaviors found in this study indicate the significance of scale, suggesting the need for further soil scaling studies.

ACKNOWLEDGEMENTS

This project is supported by the Pedometrics, Landscape Analysis, and GIS Laboratory, Soil and Water Sciences Department, University of Florida, USA. We thank the technical staff members of the Environmental Pedology Laboratory, Soil and Water Sciences Department, University of Florida, USA for assistance with the soil carbon analysis.

REFERENCES

Addiscott T, Smith J, Bradbury N. 1995. Critical evaluation of models and their parameters. *J Environ Qual.* **24**: 803–807.

Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, Roger J M, McBratney A B. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trac Trend Anal Chem.* **29**: 1073–1081.

Blöschl G, Sivapalan M. 1995. Scale issues in hydrological modelling: A review. *Hydrol Process.* **9**: 251–290.

Booth T H, Nix H A, Hutchinson M F, Busby J R. 1987. Grid matching: A new method for homoclime analysis. *Agr Forest Meteorol.* **39**: 241–255.

Breiman L. 2001. Random forests. *Mach Learn.* **45**: 5–32.

Brown D J, Bricklemyer R S, Miller P R. 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma.* **129**: 251–267.

Chaudhary V P, Sudduth K A, Kitchen N R, Kremer R J. 2012. Reflectance spectroscopy detects management and landscape differences in soil carbon and nitrogen. *Soil Sci Soc Am J.* **76**: 597–606.

Clark R N. 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy. *In* Rencz A N (ed). Manual of Remote Sensing. Volume 3. Remote Sensing for the Earth Sciences. John Wiley and Sons, New York. pp. 3–58.

Davies A M C, Fearn T. 2006. Back to basics: Calibration statistics. *Spectr Eur.* **18**: 31–32.

Dotto A C, Dalmolin R S D, Grunwald S, ten Caten A, Filho W P. 2017. Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. *Soil Till Res.* **172**: 59–68.

Florida Fish and Wildlife Conservation Commission (FFWCC). 2003. Florida vegetation and land cover data derived from 2003 Landsat ETM+ imagery. Available online at http://myfwc.com/research/gis/data-maps/terrestrial/fl-vegetation-land-cover/ (verified on March 1, 2017).

Galvao L S, Vitorello I. 1998. Role of organic matter in obliterating the effects of iron on spectral reflectance and colour of Brazilian tropical soils. *Int J Remote Sens.* **19**: 1969–1979.

Garthwaite P H. 1994. An interpretation of partial least squares. *J Am Stat Ass.* **89**: 122–127.

Gower J C. 1971. A general coefficient of similarity and some of its properties. *Biometrics.* **27**: 857–871.

Grunwald S, Thompson J A, Boettinger J L. 2011. Digital soil mapping and modeling at continental scales: Finding solutions for global issues. *Soil Sci Soc Am J.* **75**: 1201–1213.

Guerrero C, Stenberg B, Wetterlind J, Viscarra Rossel R A, Maestre F T, Mouazen A M, Zornoza R, Ruiz-Sinoga J D, Kuang B. 2014. Assessment of soil organic carbon at local scale with spiked NIR calibrations: Effects of selection and extra-weighting on the spiking subset. *Eur J Soil Sci.* **65**: 248–263.

Guerrero C, Wetterlind J, Stenberg B, Mouazen A M, Gabarrón-Galeote M A, Ruiz-Sinoga J D, Zornoza R, Viscarra Rossel R A. 2016. Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy? *Soil Till Res.* **155**: 501–509.

Hernández N, Kiralj R, Ferreira M M C, Talavera I. 2009. Critical comparative analysis, validation and interpretation of SVM and PLS regression models in a QSAR study on HIV-1 protease inhibitors. *Chemometr Intell Lab Syst.* **98**: 65–77.

Hook P B, Burke I C. 2000. Biogeochemistry in a shortgrass landscape: Control by topography, soil texture, and microclimate. *Ecology.* **81**: 2686–2703.

Hunt G R. 1989. Spectroscopic properties of rocks and minerals. *In* Carmichael R S (ed.) Practical Handbook of Physical Properties of Rocks and Minerals. CRC Press, Taylor & Francis, Boca Raton. pp. 295–385.

John B, Yamashita T, Ludwig B, Flessa H. 2005. Storage of organic carbon in aggregate and density fractions of silty soils under different types of land use. *Geoderma.* **128**: 63–79.

Lagacherie P, Baret F, Feret J B, Madeira Netto J, Robbez-Masson J M. 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sens Environ.* **112**: 825–835.

Lark R M. 2011. Spatially nested sampling schemes for spatial variance components: Scope for their optimization. *Comput Geosci.* **37**: 1633–1641.

Lobell D B, Asner G P. 2002. Moisture effects on soil reflectance. *Soil Sci Soc Am J.* **66**: 722–727.

Mallavan B P, Minasny B, McBratney A B. 2010. Homosoil, a methodology for quantitative extrapolation of soil information across the globe. *In* Boettinger J L, Howell D W, Moore A C, Hartemink A E, Kienast-Brown S (eds.) Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Springer, Dordrecht. pp. 137–150.

Malone B P, McBratney A B, Minasny B. 2013. Spatial scaling for digital soil mapping. *Soil Sci Soc Am J.* **77**: 890–902.

Martens H, Næs T. 1989. Multivariate Calibration. John Wiley & Sons, New York.

McBratney A B. 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. *Nutr Cycl Agroecosys.* **50**: 51–62.

McBratney A B, Mendonça Santos M L, Minasny B. 2003. On digital soil mapping. *Geoderma.* **117**: 3–52.

McCarty G W, Reeves J B, Reeves V B, Follett R F, Kimble J M. 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci Soc Am J.* **66**: 640–646.

McDowell M L, Bruland G L, Deenik J L, Grunwald S. 2012a. Effects of subsetting by carbon content, soil order, and spectral classification on prediction of soil total carbon with diffuse reflectance spectroscopy. *Appl Environ Soil Sci.* **2012**: 294121.

McDowell M L, Bruland G L, Deenik J L, Grunwald S, Knox N M. 2012b. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. *Geoderma.* **189**–**190**: 312–320.

Minasny B, McBratney A B. 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma.* **142**: 285–293.

Minasny B, McBratney A B. 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemometr Intell Lab Syst.* **94**: 72–79.

Minasny B, Tranter G, McBratney A B, Brough D M, Murphy B W. 2009. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma.* **153**: 155–162.

Minasny B, McBratney A B. 2010. Methodologies for global soil mapping. *In* Boettinger J L, Howell D W, Moore A C, Hartemink A E, Kienast-Brown S (eds.) Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Springer, Dordrecht. pp. 429–436.

Morellos A, Pantazi X E, Moshou D, Alexandridis T, Whetton R, Tziotzios G, Wiebensohn J, Bill R, Mouazen A M. 2016. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosyst Eng.* **152**: 104–116.

Natural Resources Conservation Service (NRCS), U.S. Department of Agriculture (USDA-NRCS). 2009. Soil survey geographic database (SSURGO). Available online at http://soils.usda.gov/survey/geography/ssurgo/ (verified on July 3, 2017).

Nocita M, Kooistra L, Bachmann M, Müller A, Powell M, Weel S. 2011. Predictions of soil surface and topsoil organic carbon content through the use of laboratory and field spectroscopy in the Albany Thicket Biome of Eastern Cape Province of South Africa. *Geoderma.* **167**–**168**: 295–302.

Peng X T, Shi T Z, Song A H, Chen Y Y, Gao W X. 2014. Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods. *Remote Sens.* **6**: 2699–2717.

Rees R M, Bingham I J, Baddeley J A, Watson C A. 2005. The role of plants and land management in sequestering soil carbon in temperate arable and grassland ecosystems. *Geoderma.* **128**: 130–154.

Reeves III J, McCarty G, Mimmo T. 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. *Environ Pollut.* **116**: S277–S284.

Sankey J B, Brown D J, Bernard M L, Lawrence R L. 2008. Comparing local *vs.* global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. *Geoderma.* **148**: 149–158.

Sarkhot D V, Grunwald S, Ge Y, Morgan C L S. 2011. Comparison and detection of total and available soil carbon fractions using visible/near infrared diffuse reflectance spectroscopy. *Geoderma.* **164**: 22–32.

Savitzky A, Golay M J E. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem.* **36**: 1627–1639.

Shepherd K D, Walsh M G. 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Sci Soc Am J.* **66**: 988–998.

Thissen U, Pepers M, Üstün B, Melssen W J, Buydens L M C. 2004. Comparing support vector machines to PLS for spectral regression applications. *Chemometr Intell Lab Syst.* **73**: 169–179.

Turner M G, O'Neill R V, Gardner R H, Milne B T. 1989. Effects of changing spatial scale on the analysis of landscape pattern. *Landscape Ecol.* **3**: 153–162.

United States Geological Survey (USGS). 1999. National elevation dataset (NED). Available online at http://ned.usgs.gov/ (verified on July 3, 2017).

Vapnik V. 2000. The Nature of Statistical Learning Theory. 2nd Edn. Springer, New York.

Vasques G M, Grunwald S, Sickman J O. 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma.* **146**: 14–25.

Vasques G M, Grunwald S, Sickman J O. 2009. Modeling of soil organic carbon fractions using visible-near-infrared spectroscopy. *Soil Sci Soc Am J.* **73**: 176–184.

Vasques G M, Grunwald S, Harris W G. 2010. Spectroscopic models of soil organic carbon in Florida, USA. *J Environ Qual.* **39**: 923–934.

Vasques G M, Grunwald S, Myers D B. 2012a. Influence of the spatial extent and resolution of input data on soil carbon models in Florida, USA. *J Geohys Res.* **117**: G04004.

Vasques G M, Grunwald S, Myers D B. 2012b. Associations between soil carbon and ecological landscape variables at escalating spatial scales in Florida, USA. *Landscape Ecol.* **27**: 355–367.

Viscarra Rossel R A, Walvoort D J J, McBratney A B, Janik L J, Skjemstad J O. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma.* **131**: 59–75.

Viscarra Rossel R A, Behrens T, Ben-Dor E, Brown D J, Demattê J A M, Shepherd K D, Shi Z, Stenberg B, Stevens A, Adamchuk V, Aïchi H, Barthès B G, Bartholomeus H M, Bayer A D, Bernoux M, Böttcher K, Brodský L, Du C W, Chappell A, Fouad Y, Genot V, Gomez C, Grunwald S, Gubler A, Guerrero C, Hedley C B, Knadel M, Morrás H J M, Nocita M, Ramirez-Lopez L, Roudier P, Campos E M R, Sanborn P, Sellitto V M, Sudduth K A, Rawlins B G, Walter C, Winowiecki L A, Hong S Y, Ji W. 2016. A global spectral library to characterize the world's soil. *Earth Sci Rev.* **155**: 198–230.

Webster R, Oliver M A. 2007. Geostatistics for Environmental Scientists. 2nd Edn. John Wiley & Sons, Chichester.

Wetterlind J, Stenberg B. 2010. Near-infrared spectroscopy for within-field soil characterization: Small local calibrations compared with national libraries spiked with local samples. *Eur J Soil Sci.* **61**: 823–843.

Williams P C. 1987. Variables affecting near infrared reflectance spectroscopic analysis. *In* Williams P C, Norris K H (eds.) Near Infrared Technology in the Agricultural and Food In-

dustries. American Association of Cereal Chemists, St. Paul. pp. 143–167.

Wu G J, Jones K B, Li H B, Loucks O L. 2006. Scaling and Uncertainty Analysis in Ecology: Methods and Applications. Springer, Dordrecht. pp. 143–167.

Xiong X, Grunwald S, Corstanje R, Yu C, Bliznyuk N. 2016. Scale-dependent variability of soil organic carbon coupled to land use and land cover. *Soil Till Res.* **160**: 101–109.

Yimer F, Ledin S, Abdelkadir A. 2006. Soil organic carbon and total nitrogen stocks as affected by topographic aspect and vegetation in the Bale Mountains, Ethiopia. *Geoderma.* **135**: 335–344.