Transferability and scaling of soil total carbon prediction models in Florida

The applicability, transfer, and scalability of visible/near-infrared (VNIR)-derived soil models are still poorly understood. The objectives of this study in Florida, U.S. were to: (i) compare three methods to predict soil total carbon (TC) using five fields (local scale) and a pooled (regional scale) VNIR spectral dataset, (ii) assess the model's transferability among fields, and (iii) evaluate the up- and down-scaling behavior of TC prediction models. A total of 560 TC-spectral sets were modeled by Partial Least Square Regression (PLSR), Support Vector Machine (SVM), and Random Forest. The transferability and up- and down-scaling of 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 square prediction error (RMSE) = 0.09%, residual predication deviation (RPD) > 2.70% , and ratio of prediction error to inter-quartile range (RPIQ) > 4.54. PLSR performed substantially better than SVM to scale and transfer models. Upscaled soil TC models performed somewhat better in terms of model fit (R^2), RPD, and RPIQ, whereas downscaled models showed less bias and smaller RMSE based on PLSR. Given the many factors that can impinge on empirically derived soil spectral prediction models, as demonstrated by this study, more focus on the applicability and scaling of them is needed.

Transferability and scaling of soil total carbon prediction models in Florida

Sabine Grunwald[1*], Congrong Yu[1,2] and Xiong Xiong[1]

[1] Department of Soil and Water Science, University of Florida, Gainesville, FL 32611, USA

[2] College of Hydrology and Water Resource, Hohai University, Nanjing, Jiangsu 210098, China

[*] Corresponding author: Sabine Grunwald, 2181 McCarty Hall A, P.O. Box 110290, Gainesville, FL 32611-0290 USA. Phone: 352-294-3145. Email: sabgru@ufl.edu

1 Keywords: Soil total carbon, visible-near infrared spectroscopy, model transferability, scaling

2 Abbreviation: DRS, diffuse reflectance spectroscopy; MIR, mid-infrared; PLSR, partial least
3 square regression; RMSE, root mean squared error; RPD, residual predication deviation; RPIQ,
4 ratio of prediction error to inter-quartile range; RF, random forest; SOC, soil organic carbon;
5 SVM, support vector machine; TC, soil total carbon; VNIR, visible/near-infrared.

1. 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. But the cost and time involved in 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 for TC and carbon fractions (McCarty et al., 2002; Reeves III, McCarty and Mimmo, 2002; Viscarra Rossel et al., 2006; Vasques, Grunwald and Sickman, 2009; Vasques, Grunwald and Harris, 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. Spectral soil carbon models are poised to contribute to spatially explicit regional and global carbon assessment and monitoring of soil carbon evolution. However, 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 information through explicit mathematical expressions and statistical relationships (scaling equations) or process-based simulation (Blöschl and Sivapalan, 1995; Wu et al., 2006). 'Scaling up' (or up-scaling) translates information from finer scales (smaller grain sizes or extents) to broader scales (larger grain sizes or extents), whereas 'scaling down' (or down-scaling) translates information from broader scales 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 the coarser scale system behaves like the average finer scale system, which implies that processes are linear. Non-linear 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 impacting the 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' may 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, Minasny and McBratney (2010) asserted that the more similar regions are in terms of soil-environmental properties the more likely it is to successfully transfer a soil prediction model. Several methods have been utilized to assess the similarity among soil ecosystems, among them the Gower similarity index (Gower, 1971) and soil taxonomic distance metrics (Minasny and McBratney, 2007).

A review of spatial scaling concepts and procedures used in digital soil mapping (DSM) was provided by Malone, McBratney and Minasny (2012). Vasques, Grunwald and Myers (2012a) assessed the scaling effects of soil carbon models considering the geographic extent and grain size of models in Florida, U.S. Vasques, Grunwald and Myers (2012b) investigated multi-

scale behavior of soil carbon and identified those environmental factors that imposed most effects on the predictions of soil carbon at three nested spatial scales. But scaling studies of VNIR soil prediction models are rarely found in the literature. For example, Brown, Bricklemyer et al. (2005) found that about half of the transferred VNIR soil carbon models did not perform well in Montana, U.S. 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 severely degraded when applied to other regions in eastern Australia. They asserted that local calibration of spectral models is preferable to avoid the measurement biases between laboratories in different regions. In addition, they argued that the calibrated spectral-soil functions are specific to the soil types in an area limiting their transferability to other application sites.

Shepherd and Walsh (2002) initiated the idea of 'ubiquitous' spectral libraries to predict soil properties. After once developed (calibration phase) and validated these spectral soil libraries can be applied elsewhere, similar to pedo-transfer functions (PTF). Brown, Bricklemyer and Miller (2005) presented such an 'ubiquitous' global soil spectral library derived from VNIR spectra using a large soil dataset of 3768 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 on samples ranging from 0.0 to 536.8 g kg^{-1} SOC and a median of 4.7 g kg^{-1}. They suggested that VNIR soil characterization has the potential to replace or augment standard soil characterization techniques where rapid and inexpensive analysis is required. One limitation of VNIR soil carbon modeling is that predictions are relatively poor in the low soil carbon attribute domain space (McDowell et al., 2012b; Vasques, Grunwald and Harris, 2010).

There are numerous factors that potentially influence the prediction performance of transfer and scalability of spectral soil models. Among them are the (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 pedogenic processes to store or lose carbon in soils, (iii) other soil properties (such as minerals, sesquioxides, and texture) that may mask, interfere or amplify the identification of spectral predictors to infer on soil carbon, (iv) differences in measurement protocols of soil carbon and spectral data, (v) spectral instruments, (vi) statistical methods used to develop soil carbon predictions, and (vii) the (dis)similarity between soil samples used to develop the spectral prediction model and those that are used to be estimated. Given the multitude of potential factors that may impact 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 down-scaling behavior.

The specific objectives were to (i) compare the performance of three modeling types to predict soil TC using five fields (local) and a pooled (regional) VNIR spectral dataset, (ii) assess the model's transferability among five representative field sites in Florida, (iii) evaluate the up-scaling behavior of TC prediction models from local (field) to regional scale, (iv) evaluate the down-scaling behavior of TC prediction models from regional to local scale, and (v) examine the constraining factors in model transferability and scaling.

2. Materials and Methods

2.1. Study Area

Five fields (each of size ~0.25 km^2) were selected that represent prominent soil-landuse types in Florida, U.S. (see Xiong, 2013). Table 1 provides a description of the main landscape characteristics of each field.

2.2. Field Sampling

Soil samples were collected in each of the five fields with the same unbalanced nested spatial sampling design (Lark, 2011) as described by Xiong (2013). 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 point 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).

2.3. Laboratory Analysis

Soil total carbon content was measured by 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, then ground and passed through a 2-mm sieve. Around 50 g of each sample was ball-milled for 3 minutes, from which ~500 mg was combusted at 900°C for about 3 minutes on a Shimadzu TOC-5050 analyzer to measure the TC. There are miniscule amounts of inorganic carbon found in Florida soils and soil organic carbon constitutes TC at many sites (Vasques, Grunwald and Harris, 2010). Since soil moisture can impact the soil spectral reflectance (Lobell and Asner, 2002), the ball-milled samples were oven dried at 40-45°C for 12 h before scanning. After cooling for 1h, they were scanned using the QualitySpec Pro Spectroradiometer (Analytical Spectral Devices Inc., Boulder, CO) 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, NH). An

127 average reflectance spectral curve was obtained for each sample for further transformation and
128 modeling.

129 **2.4. Pre-processing Transformations**

130 The laboratory measured TC in the five fields was positively skewed (Table 2). In order to
131 reduce the skewness of the TC distributions and the model standard error of prediction (Bellon-
132 Maurel et al., 2010), the TC data were logarithm-transformed to approximate normal
133 distributions. Before constructing models, two pre-processing transformations were applied to the
134 soil reflectance curves. For all VNIR spectra, we used the pre-processing methods that worked
135 best in a previous study in Florida (Vasques, Grunwald and Sickman, 2008). First, the
136 reflectance curves were smoothed across a moving window of nine nm using the Savitzky-Golay
137 algorithm with a third-order polynomial to reduce the random noise (Savitzky and Golay, 1964).
138 Second, the first-degree Savitzky-Golay derivative, with a search window of seven measurements
139 and second-order polynomial, was applied to the smoothed curves.

140 **2.5. Regression Techniques**

141 In order to compare the predictive performance of the three different VNIR diffuse reflectance
142 models to predict TC in Florida, the Whole TC dataset (n=560) was randomly split into
143 calibration set (CAL) (70%) with n=392 and validation set (VAL) (30%) with n=168. Three
144 different multivariate regression techniques were applied to develop spectral models that were
145 consequently evaluated to predict soil TC using the VNIR spectra: Partial Least Square
146 Regression (PLSR) (Martens and Næs, 1989), Support Vector Machine (SVM) (Vapnik, 2000),
147 and Random Forest (RF) (Breiman, 2001). The Partial Least Square Regression approach is well
148 suited for the prediction of regression models with a large number of highly collinear predictor
149 variables (Garthwaite, 1994). In PLSR, the target variable (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 as much as possible of the 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 calibration set. In this paper, 15 PCs were employed, which represented over 90% of the soil samples' variation. Support Vector Machine was originally proposed by Vapnik (2000) within the area of statistical learning theory. To address non-linearity in input-output data relationships, SVM 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' method was performed to find the best choices for parameters 'Cost' and 'Sigma' for the kernel. Random Forest is a learning ensemble consisting of a bagging of un-pruned decision tree learners, with randomized selection of predictor variables at each split (Breiman, 2001). All three regression methods were employed to relate spectral data to the logTC transformed data. First, leave-one-out (LOO) cross-validation was employed to evaluate the model performance of the CAL datasets. Second, independent validation was used to assess the model performance using the VAL datasets.

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

2.6. Model Transferability and Scaling 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

174 relationships (PLSR) and one that models non-linear relationships (SVM). Since the behavior of
175 SVM and RF were very similar in terms of transferability and scalability this paper only presents
176 the results derived from SVM.

177 In this study, the definitions of 'model transfer', 'scale transformation' and 'up-/down-
178 scaling' as provided by Turner et al. (1989), Wu et al. (2006), and Blöschl and Sivapalan (1995)
179 were adopted. Hence, 'transferability' denotes the transfer (or application) of a VNIR-based soil
180 TC prediction model (Models 1 to 5) developed (calibrated) and validated at one field site (Fields
181 1, 2, 3, 4, and 5, respectively) to another field site (Fields 1, 2, 3, 4, and 5, respectively) (Fig. 1-
182 b). Note that all the field sites were of equal size and contained the same number of observations
183 to build models. This process does not entail scaling of neither grain size nor extent, which is
184 commonly denoted as 'extrapolation' (Wu et al., 2006). The five models developed in the five
185 fields were denoted as Model 1 to Model 5. The model performance at calibration sites was
186 assessed using LOO cross-validation reporting R^2 and RMSE and transferability was assessed
187 using R^2, RMSE, RPD, and RPIQ.

188 In this paper, 'scalability' denotes a change in the extent (size) of the geographic area
189 represented by models, 'up-scaling' refers to an escalation of the area (i.e., from smaller to larger
190 extent), and 'down-scaling' refers to the contraction of the area (i.e., from larger to smaller
191 extent) (after Wu et al., 2006) (Fig. 1-c and 1-d). To assess the down-scaling behavior, the
192 regional SUB-W models (representing the pooled field areal coverage of ~1.25 km^2) were applied
193 to each of the five fields (each ~0.25 km^2 in size) (Fig. 1-c). And vice versa, to assess the up-
194 scaling performance, the TC models using PLSR and SVM developed for each of the five fields
195 were applied to the regional SUB-W dataset (Fig. 1-d). To exclude the impact of observation size
196 from the scaling procedure the observation size was kept constant at n = 112 for SUB-W and
197 each of the five field models during the scaling procedure. The same error statistics as outlined
198 above were used to evaluate scaling behavior of TC models.

In the scaling analysis, a pooled subset-whole (SUB-W) dataset was created (with n = 112) from the five field dataset (with n = 560). The observation size of the SUB-W 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 x 5 = 560) field set with 1/5 random samples contributed from each of the five fields. The selection process of the SUB-W dataset is shown in Fig. 1-a. The models calibrated by the spectral SUB-W data were evaluated using LOO cross-validation.

2.7. Similarity of Soil-environmental Conditions among Fields and across Scales

To examine the constraining effect of soil-environmental conditions on the transfer of soil spectral models across fields and scales similarities among fields and across scales were characterized by the SCORPAN factors (McBratney, Mendonça Santos and Minasny, 2003). The SCORPAN factors include soil characteristics (S), climate (C), organism, vegetation, land use (O), relief (R), parent material (P), age (A), and space (N). Assuming homology of soil-forming factors between a calibration area and the region of interest, Mallavan, Minasny and McBratney (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 in Mallavan, Minasny and McBratney (2010), was employed to measure the similarity in soil-forming factors among fields according to Eq. (1). Important variables that were included in the similarity analysis are shown in Table 1.

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

where S_{ij} is the Gower similarity coefficient between two 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. The Gower similarity coefficients were compared with models' prediction performance.

3. Results

3.1. Descriptive Statistics

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

The soil TC mean and median (%) was highest in Fields 2 and 5 and declined in Fields 4, 3, and 1 (Table 2) resembling a hydrologic gradient as indicated by the available water capacity (AWC) with highest values in Fields 4 and 2, and declining in Fields 5, 3, and 1 (Table 1). The trajectory of soil TC values mirrors also the land use / land cover gradient with highest soil TC found under Mesic Upland Forest and Rangeland and lowest soil TC occurring in Xeric Upland Forest.

3.2. Assessment of Prediction Performance for Soil Total Carbon

The results of predicting soil logTC in calibration and validation modes using three different methods are shown in Table 3. The performance of SVM and RF models was very similar in both LOO cross-validation and validation modes. In LOO cross-validation, the PLSR models performed slightly better, with the highest R^2 (0.88) and lowest RMSE (0.08%) 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 validation mode, the PLSR models had a smaller bias, although the R^2 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 values aligned close to the 1:1 line (Fig. 2). However, the SVM and RF models tended to slightly over-predict when logTC was smaller than zero, and under-predict in the high logTC data range. The points of the PLSR models were almost homogenously distributed along the 1:1 line. The high RPD ($\geq$ 2.70), RPIQ ($\geq$ 4.54) values for all three models confirmed their excellent prediction accuracies.

3.3. Transferability and Scaling Analysis

The TC prediction performances declined at varying degrees when the PLSR and SVM models were transferred to the field outside the calibration geographical area (Fig. 1-b), downscaled (Fig. 1-c), and upscaled (Fig. 1-d). Overall, the PLSR model showed better down-scaling and up-scaling performances than those derived from SVM. In the SVM model set, only down-scaling produced acceptable TC predictions. Although the SUB-W (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 (see Table 4) compared to the calibration models (with n = 392) in Table 3, the Model SUB-W could still effectively represent the VNIR-TC relationship at the regional scale. The goodness-of-fit statistics for Models 1 to 5 at field scale ranged from 0.46 to 0.69 R^2 (PLSR) and 0.33 to 0.59 R^2 (SVM) (Table 4), indicating that the performance of models developed at field scale were not as good as the regional scale model using the same size of calibration sample (n = 112).

3.3.1. Transferability of Soil Carbon Prediction Models among Field Sites

In order to test the model transferability at field scale PLSR and SVM models developed and cross-validated at one field were applied to the other four fields. Results of the transferability

of PLSR models are summarized in Table 5 and the SVM models in Table 6. The R^2 values of PLSR models were below 0.39, except when Model 1 was applied to predict the TC of Field 2 (R^2 = 0.53) and Field 3 (R^2 = 0.51). The predicted TC of Field 1 by Models 3, 4, and 5 were severely biased (0.37%). The high bias and RMSE values with low RPD and RPIQ values suggested the transferability among field sites was severely constrained using PLSR models. The R^2 values derived from the SVM models were much lower than those derived from PLSR models, with several values even smaller than 0.01, indicating that the SVM model transferability among field sites was severely constrained (Table 6). The limited transferability using SVM was confirmed by high bias and prediction errors (Table 6). Interestingly, when two models calibrated in different fields were transferred to each other, the models behaved in different ways. For example, although Model 1 predicted the TC in Field 2 well with R^2 0.53, the performance of Model 2 was poor when transferred to Field 1.

3.3.2. Down-scaling of Soil Carbon Prediction Models

To test the down-scaling performance the regional pooled model (SUB-W) was applied to predict the TC of the five fields (Tables 7 and 8). Both of the PLSR and SVM models successfully predicted the TC of Field 2 and Field 4 with R^2 larger than 0.47, but the performance in Field 1 and Field 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 = 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 down-scaling behaviors were acceptable, but far from stellar performance of models. Important to note is that down-scaling behavior from regional to field scale differed substantially among fields.

3.3.3. Up-scaling of Soil Carbon Prediction Models

290 Tables 9 and 10 show the up-scaling behavior of the five field models (Models 1 to 5) to
291 predict the TC of SUB-W at the regional scale. The performance of the five PLSR field scale
292 models, except for Model 3, was as good as the downscaled models using PLSR. The R^2 of
293 upscaled PLSR Models 1, 2, and 4 were all above 0.50 and the RPIQ values higher than 1.84,
294 suggesting moderately good up-scaling performance (Table 9). The PLSR Model 3 failed to
295 predict TC of SUB-W, which was due to its poor performance in LOO cross-validation mode
296 (Table 4). Although the PLSR Models 1, 2, 4, and 5 had high R^2 values, the large bias and RMSE
297 values showed that up-scaling did not perform as well as down-scaling. In contrast, the up-
298 scaling of SVM models was very poor. The field scale SVM models predicted the TC of the
299 SUB-W dataset with R^2 below 0.19 and RMSE higher than 0.23%, and RPIQ values were slightly
300 lower than that of PLSR models.

301 **3.4. Gower Similarity Coefficient**

302 The Gower similarity coefficients were all above 0.50 (Table 11). The coefficients
303 between SUB-W dataset and each of the five fields were higher than those between the five
304 fields. Correspondingly, down-scaling performance of the SUB-W models using PLSR and SVM
305 outperformed the field models when transferred to other fields.

306 **4. Discussion**

307 **4.1. Prediction Performance of Spectral Prediction Models**

308 The TC predictions derived from all three multivariate methods (PLSR, SVM, and RF) at
309 the five field sites showed good performance within the observation range of 0.31 to 3.55 %
310 which was slightly narrower than in other studies (McCarty et al., 2002; Vasques et al., 2008;
311 Sarkhot et al., 2011). Brown, Bricklemyer and Miller (2005) found that VNIR models developed
312 using boosted regression trees (BRT) outperformed PLSR to predict SOC and soil TC, while

McDowell et al. (2012b) found no significant difference among PLSR and RF ensemble regression trees to predict soil TC on Hawaiian soils. Minasny and McBratney (2008) and Minasny et al. (2009) in Australia found excellent predictions for SOC and TC using regression rules (Cubist approach). In contrast, Vasques, Grunwald and Harris (2010) identified SOC predictions made by 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 may outperform several others to make SOC or TC predictions from VNIR spectra.

4.2. Factors that impact the Transferability and Scalability of Prediction Models

Overall, PLSR models performed better to transfer and scale than SVM models. This implies that linear relationships between VNIR spectra and soil TC (quantified by PLSR) were more pronounced than non-linear, complex relationships (quantified by SVM). Reasons that constrain the transferability and scaling of soil prediction models may be explained by differences in the: (i) spectral data domain space, (ii) soil attribute domain space, (iii) methods that determine the internal model structure of VNIR-TC relationships, and (iv) environmental domain space of 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 allowed comparing models from a statistical perspective. One factor that may have impacted the up-scaling and down-scaling behavior of TC models is the sampling density that differed among the field sites (each ~0.25 km^2 with a density of observations of ~448 per km^2) and the SUB-W regional set (representing an area of ~1 km^2 with a density of observations of ~112 per km^2). In essence, the sample number of the field model and regional model was the same (n = 112), while the geographical area of the regional model was five times larger than each of the field models. According to McBratney et al. (1998) up-scaling 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, Grunwald and Myers (2012b) who scaled soil TC from small region (5.6 km^2), intermediate region (3,500 km^2) to coarse scales (150,000 km^2) in Florida.

4.2.1. Spectral Data Domain Space

The transferability and scaling of models may be also 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, Grunwald and Sickman (2008; 2009) and Vasques, Grunwald and Harris (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 mid-infrared regions (Brown, Bricklemyer and Miller, 2005). The features associated with TC can be masked or distorted by Fe-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 may mask or interfere with the prediction of a given property of interest (e.g., soil TC); and thus, impact the transferability of models. In this study the soil texture differed only slightly among the five sites with sand content ranging between 90.8 to 98.6% and clay content between 1.2 to 5.2%. Hence, the effect of soil texture imposed on 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 differed among sites (Table 1), with Entisols (Psamments), Ultisols (Aquults, Udults), Inceptisols (Udepts), and Spodosols (Aquods), suggesting that the mineralogy, sesquioxides, and other chemical and physical soil properties differed substantially among sites. This may have constrained the transfer and scalability of VNIR-based TC prediction models due to masking or distortion effects in the spectral data domain.

4.2.2. Soil Attribute Domain Space

The soil attribute space, i.e., the upper and lower bounds and dispersion of soil TC used to build spectral-based prediction models (Table 2), may 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 soil TC values of CAL, VAL, and the SUB-W sets matched reasonably well the minimum of 0.31% and maximum of 3.55% of the Whole data set. However, the differences in soil TC among field sites were profound (Table 2). The transferability of a TC prediction model to other sites may lead to an extrapolation outside the soil attribute observation range of the original field data which may impacts 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 set. Brown et al. (2005) demonstrated the implications of spectral-based model transfer to predict soil carbon in other fields in Montana, U.S. where the SOC values differed widely among field sites (minimum of 1.93 g kg^{-1} to maximum of 15.82 g kg^{-1}). In their study 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 in turn 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 pre-screening for spectral similarity.

In this study Models 1, 3, and 4, that resembled the TC range of SUB-W most closely with TC minimum of 0.32% and TC maximum of 2.85%, did not show persistent responses in terms of transferability based on PLSR (Table 5). For example, Model 3 (developed in Pineland and Psamments) failed to transfer well to Field 4, whereas the opposite was found for the transfer behavior of Model 4 (developed in Improved Pasture and Udults) to Field 3. The models that exceeded the upper bound TC of SUB-W (Model 2 developed in Mesic Upland Forest and

diverse soil suborders; and Model 5 developed in Rangeland and Aquods) transferred somewhat equally well to other fields, suggesting that the attribute domain range did not substantially impact model transferability. These findings were confounded in down-scaling mode. The SUB-W Model (with min. TC = 0.32% and max. TC = 2.85) degraded substantially more when applied to Field 5 with a mismatched attribute domain space (with min. TC = 1.02% and max. TC = 3.55%) than to field sites with similar attribute domain space (e.g., Field 4 with min. TC = 0.56% and max. = 2.84%) or field sites with narrower attribute domain space (e.g., Field 1 with min. TC = 0.32% and max. = 1.12%) (Tables 2 and 7). In up-scaling 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 2 and 9). These findings substantiate that no clear conclusions emerge that link 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 also the internal variability (variance) of soil attributes that potentially impacts behavior. Addiscott, Smith and Bradbury (1995) pointed out that an increase in parameter variance may cause problems by interacting with the linearity / non-linearity in the process represented by the model. McBratney (1998) and Grunwald et al. (2011) asserted that an increase in the variance of soil attributes can impact 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% in SUB-W and 0.55% in Whole, respectively). The low CV in Field 5 limited somewhat the transferability of TC models to other field sites, more so than the transferability of other models to field sites (Table 5). It is interesting to note that Model 3 (developed in Pineland and Psamments), which had the highest variability in TC among field sites, performed poorest in terms of transferability to other

field sites. In contrast, the SUB-W model with higher variability in TC than Model 3 performed substantially better suggesting that the internal variability in attributes are one but perhaps not the most controlling factor that limits model transferability and scalability. The down-scaling performance of the SUV-W model with the highest CV of 0.53% had less impact on Field 3 (with intermediate CV of 0.42%) but substantially degraded model performance when applied to Field 5 (with the lowest CV of 0.26%) for PLSR and SVM models (Tables 7 and 8). These findings suggest that although the SUV-W model represented the variability in TC of the five fields it severely degraded the down-scaling to those fields that were more homogeneous in soil TC (e.g., Field 5). Up-scaling of Model 3 to SUB-W failed and was severely muted for Model 5, whereas Models 1, 2, and 4 were less impacted in terms of model performance. This confirms the assertion that the variance in TC plays a role in the scalability of models; however, there are other substantial factors that confound findings.

Spiking of a spectral model with local samples has been suggested to improve soil predictions (Sankey et al., 2008; Wetterlind and Stenberg, 2010) though its success is highly dependent on multiple factors including the ratio between the 'number of spike samples' and 'number of samples in the spectral library', characteristics of the soil attribute and spectral domain spaces, and the methods used to develop spectral-based soil prediction models. Although the same constraints, mechanisms, and effects impact spiking and scaling of chemometric models, the aims are inherently different. Spiking aims to stabilize/improve predictions of soil properties by adding more observations to the dataset, whereas scaling aims to understand the factors and processes impacting the scaling behavior.

4.2.3 Methods (Model Types)

Regression methods use different strategies to relate predictors (here: spectral data) and a response variable (here: soil TC). The underlying strategies for predictor selection are different

for PLSR and SVM as described in the methods section impacting transfer and scale responses. If the internal model structure that describes the relationship between spectral predictors and soil TC is not stable when it is scaled, it suggests scale variant behavior. As expected, the PLSR (Fig. 3-a) and SVM (Fig. 3-b) models showed differences in the selection of spectral predictors in Models 1 to 5 and the SUB-W Model. Thissen et al. (2004) has 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 physico-chemical composition of the soil samples differs. In this study PLSR was more robust than SVM to transfer models among sites. The PLSR models (Fig. 3-a) mainly focused on three regions to identify spectral predictors: ~350 nm, ~1860 nm and ~2200 nm, which represented the reflection region of organic matter (Galvao and Vitorello, 1998); $O\text{-}H^{-}$, water, C-H, C-N, C-O, N-H (Vasques, Grunwald and Sickman, 2008); and calcium carbonate (2206 nm and 2341 nm) (Lagacherie et al., 2008), M-OH, and various C-O (Brown, Bricklemyer and Miller, 2005). On the other hand, the top 50 important spectral wavelengths of the SVM models (Fig. 3-b) were found around ~670 nm, ~1400 nm, ~1800 nm and ~2200 nm. In particular, 670 nm is indicative of iron oxide features (McDowell et al., 2012b) and 1400-1900 nm are absorption regions of O-H and water showing that iron oxide and crystallization water impacted the TC predictions. In the VNIR spectral range only overtones are mapped, which differs from other spectral methods, such as mid-infrared sensing, that more directly respond to the chemical composition of samples (McDowell et al., 2012a; b).

Although SVM is advantageous to model complex, high-dimensional spectral datasets because it can model nonlinear structures it performed poorly to transfer and scale models. This can be explained by the high susceptibility of SVM to overfitting (Hernández et al., 2009). The substantially larger amount of spectral values selected as important in the SVM model compared to the PLSR model suggests overfitting (Fig. 4). In the SVM model most of the spectral predictors were assigned large values compared with the PLSR and RF models (Fig. 4). In

addition, non-linear relationships between spectral data and TC may be rather weak to limit its competitiveness when compared to a more robust, linear method such as PLSR (Hernández et al., 2009).

4.2.4. 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, Grunwald and Myers, 2012b), and topography (Yimer, Ledin and Abdelkadir, 2006). Mallavan, Minasny and McBratney (2010) argued that soil attributes correlate consistently with environmental factors assuming 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, under the condition 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 1). The homology among environmental conditions explained a substantial amount of the ability to transfer TC models to other field sites and scales in this study (Fig. 5 and Table 11). Minasny et al. (2009) found that the transfer of MIR spectral SOC prediction models among three different regions in Australia did not perform well due to differences in parent material and climate in which soils have formed in Queensland, New South Wales, and Victoria. Unfortunately no explicit similarity analysis of environmental factors was presented in their study and relationships between soil-environmental factors are not clear (Minasny et al., 2009). Although the R^2 of transferred models were still moderate all models showed significant bias. Studies that test not only for similarity in

soil TC (or other soil properties) among sites, but also consider the similarity in environmental factors that form those soil properties are still rare in the soil science literature.

5. Conclusions

This study showed that, although spectral models to predict soil TC with three different methods (PLSR, SVM, and RF) were successful in calibration and validation modes at five different fields nested within a large sand-dominated region in the U.S., the transferability and up- and down-scaling of models were limited by the following factors: (i) the spectral data domain space, (ii) soil attribute domain space, (iii) 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 impacting the transferability of models among field sites, up-scaling, and down-scaling behavior of spectral soil prediction models.

Overall, the transferability and scalability of prediction models derived from PLSR were better and more robust than those derived from SVM. But no universal trend was found indicating which of the four investigated factors (i to iv) had the most impact that constrained transferability and scalability. Interestingly, up-scaling of soil TC models performed somewhat better than down-scaled models in terms of model fit (R^2), RPD, and RPIQ, whereas down-scaled models showed less bias and smaller RMSE derived from PLSR. These findings have implications for the development of 'universal' spectral-based soil models aiming to predict soil properties for a diverse set of different soils formed in different environmental conditions covering a wide range of geographic settings, at its extreme the whole globe. Those 'universal' spectral libraries are based on the premise that soil predictions (e.g., soil TC) can be made anyplace because they are built using soil spectral datasets that characterize exhaustively the attribute feature space. This assertion is limited by the fact that a large number of interacting

506 factors of soils, spectra, and environmental properties are needed to represent the exhaustive
507 sample population which has not materialized yet.

508 Furthermore, the stationarity in mean and variance in local (field) calibrations of spectral
509 soil prediction models are usually easier to meet though can have severe effects on scale-variant
510 behavior of models at escalating spatial scales. The confounding trends in SOC up- and down-
511 scaling behavior found in this study suggests that scale matters indicating the need for further soil
512 scaling studies.

513 Findings from this study purport the idea that the selection of representative soils (Model
514 SUB-W) spanning across the attribute and spectral domains of local sites outperform predictive
515 capabilities of local models (Fields 1 to 5) (Table 4). However, this is confounded by the fact that
516 the SUB-W Model when applied to local fields substantially degraded in terms of performance
517 (Table 7). In analogy, a 'universal' spectral library may also suffer from severe degradation
518 effects predicting at local (site-specific) scale. Specifically, if 'universal' soil spectral libraries are
519 created using crawling approaches (e.g., assembling soil-spectra data from publicly available
520 databases irrespective of quality) or spiking/pooling of soil-spectral data that is random instead of
521 strategic (e.g., based on funded projects that generate data) leading to extremely unbalanced
522 datasets where one geographic region is over- and others underrepresented. Given the many
523 factors that can impinge on empirically derived soil spectral prediction models, as demonstrated
524 by this study, more focus on the applicability and scaling of them is needed. This study
525 juxtaposed local and regional predictions, transferability, and scalability of soil TC models
526 derived from VNIR spectra within a subtropical region in the southeastern U.S. The constraints
527 and limitations of soil spectral models identified in this research may also be found in other
528 regions and spectral libraries that intent to have universal applicability.

529 **Acknowledgements**

We thank the technical staff members of the Environmental Pedology Laboratory, Soil and Water Science Department 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(5): 803–807.

Bellon-Maurel V, Fernandez-Ahumada E., Palagos B, Roger JM, McBratney AB. 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(9): 1073–1081.

Blöschl G., Sivapalan M. 1995. Scale issues in hydrological modelling: A review. *Hydrol. Process.* 9(3-4): 251–290.

Booth TH, Nix HA, Hutchinson MF Busby JR. 1987. Grid matching: a new method for homoclime analysis. *Agr. Forest Meteorol.* 39(2): 241–255.

Breiman L. 2001. Random forests. *Mach. Learn.* 45(1): 5–32.

Brown DJ, Bricklemyer RS, Miller PR. 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma* 129(3–4): 251–267.

Chaudhary VP, Sudduth KA, Kitchen NR, Kremer RJ.. 2012. Reflectance spectroscopy detects management and landscape differences in soil carbon and nitrogen. *Soil Sci. Soc. Am. J.* 76(2): 597–606.

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

Davies AMC, Fearn T. 2006. Back to basics: calibration statistics. *Spectrosc. Eur.* 18: 31–32.

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

Galvao LS, 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(10): 1969–1979.

Garthwaite PH. 1994. An interpretation of Partial Least Squares. *J. Am. Stat. Ass.* 89(425): 122–127.

Ge Y, Morgan CLS, Grunwald S, Brown DJ, Sarkhot DV. 2011. Comparison of soil reflectance spectra and calibration models obtained using multiple spectrometers. *Geoderma* 161(3-4): 202–211.

Gogé F., Gomez C, Jolivet C, Joffre R. 2014. Which strategy is best to predict soil properties of a local site from a national Vis–NIR database? *Geoderma* 213: 1–9.

Gower JC. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27(4): 857–871.

Grunwald S. 2006. *Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics*. CRC Press, Taylor & Francis, Boca Raton, Florida.

Grunwald S., Thompson JA, Boettinger JL. 2011. Digital soil mapping and modeling at continental scales: Finding solutions for global issues. *Soil Sci. Soc. Am. J.* 75(4): 1201–1213.

Hernández N, Kiralj R, Ferreira MMC, 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(1): 65–77.

Hook PB, Burke IC. 2000. Biogeochemistry in a shortgrass landscape: control by topography, soil texture, and microclimate. *Ecology* 81(10): 2686–2703.

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

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(1-2): 63–79.

Lagacherie P, Baret F, Feret JB, Madeira Netto J, Robbez-Masson JM. 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sens. Environ.* 112(3): 825–835.

Lark RM. 2011. Spatially nested sampling schemes for spatial variance components: Scope for their optimization. *Comput. Geosci.* 37(10): 1633–1641.

Law BE, Turner D, Campbell J, Lefsky M, Guzy M, Sun O, Tuyl SV, Cohen W. 2006. Carbon fluxes across regions: observational constraints at multiple scales. p. 167–190. *In* Wu, J., Jones, K.B., Li, H., Loucks, O.L. (eds.), *Scaling and Uncertainty Analysis in Ecology: Methods and Applications*. Springer, Dordrecht, The Netherlands.

Lobell DB, Asner GP. 2002. Moisture effects on soil reflectance. *Soil Sci. Soc. Am. J.* 66(3): 722–727.

Mallavan BP, Minasny B, McBratney AB. 2010. Homosoil, a methodology for quantitative extrapolation of soil information across the globe. p. 137–150. *In* Boettinger, D.J.L., Howell, D.W., Moore, A.C., Hartemink, P.D.A.E., Kienast-Brown, S. (eds.), *Digital Soil Mapping. Progress in Soil Science*. Springer, The Netherlands.

Malone BP, McBratney AB, Minasny B. 2012. Spatial scaling for digital soil mapping. *Soil Sci. Soc. Am. J.* 77: 890-902.

Martens H, Næs T. 1989. *Multivariate Calibration*. 1st ed. John Wiley & Sons.

McBratney AB. 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. *Nutrient Cycling in Agroecosystems* 50(1): 51–62.

McBratney AB, Mendonça Santos ML, Minasny B. 2003. On digital soil mapping. *Geoderma* 117(1-2): 3–52.

McBratney AB, Minasny B, Cattle SR, Vervoort RW. 2002. From pedotransfer functions to soil inference systems. *Geoderma* 109(1-2): 41–73.

McCarty GW, Reeves JB, Reeves VB, Follett RF, Kimble JM. 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.* 66(2): 640–646.

McDowell ML, Bruland GL, Deenik JL, 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. Available at http://www.hindawi.com/journals/aess/2012/294121/abs/ (verified 3 February 2014).

McDowell ML, Bruland GL, Deenik JL, Grunwald S, Knox NM. 2012b. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. *Geoderma* 189–190(0): 312–320.

Minasny B, McBratney AB. 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma* 142(3–4): 285–293.

Minasny B, McBratney AB. 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemometr. Intell. Lab. Syst.* 94(1): 72–79.

Minasny B, McBratney AB. 2010. Methodologies for global soil mapping. p. 429–436. *In* Boettinger, D.J.L., Howell, D.W., Moore, A.C., Hartemink, P.D.A.E., Kienast-Brown, S. (eds.), *Digital Soil Mapping. Progress in Soil Science*. Springer, The Netherlands.

Minasny B, Tranter G, McBratney AB, Brough DM, Murphy BW. 2009. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma* 153(1–2): 155–162.

National Climatic Data Center (NCDC), National Oceanic and Atmospheric Administration (NOAA). 2008. Monthly surface data. Available at: http://www.ncdc.noaa.gov.

Natural Resources Conservation Service (NRCS), U.S. Department of Agriculture. 2009. Soil survey geographic database (SSURGO). Available at: http://soils.usda.gov/survey/geography/ssurgo/.

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(0): 295–302.

Rees RM, Bingham IJ, Baddeley JA, Watson CA. 2005. The role of plants and land management in sequestering soil carbon in temperate arable and grassland ecosystems. *Geoderma* 128(1-2): 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 JB, Brown DJ, Bernard ML, Lawrence RL. 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(2): 149–158.

Sarkhot DV, Grunwald S., Ge Y, Morgan CLM. 2011. Comparison and detection of total and available soil carbon fractions using visible/near infrared diffuse reflectance spectroscopy. *Geoderma* 164(1-2): 22–32.

Savitzky A, Golay MJE. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36(8): 1627–1639.

Shepherd KD, Walsh MG. 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Sci. Soc. Am. J.* 66(3): 988–998.

Thissen U, Pepers M, Üstün B, Melssen WJ, Buydens LMC. 2004. Comparing Support Vector Machines to PLS for spectral regression applications. *Chemometr. Intell. Lab. Syst.* 73(2): 169–179.

Trumbore SE, Czimczik CI. 2008. An uncertain future for soil carbon. *Science* 321(5895): 1455–1456.

Turner MG, O'Neill RV, Gardner RH, Milne BT. 1989. Effects of changing spatial scale on the analysis of landscape pattern. *Landscape Ecol.* 3(3-4): 153–162.

United States Geological Survey (USGS). 1999. National Elevation Dataset (NED). Available at: http://ned.usgs.gov/.

Vapnik V. 2000. *The nature of statistical learning theory*. Second. Springer, New York, USA.

Vasques GM, Grunwald S, Harris WG. 2010. Spectroscopic models of soil organic carbon in Florida, USA. *J. Environ. Qual*. 39(3): 923–934.

Vasques GM, Grunwald S, Myers DB. 2012a. Influence of the spatial extent and resolution of input data on soil carbon models in Florida, USA. *J. Geohys. Res.: Biogeosci.* 117(G4): 1–12.

Vasques GM, Grunwald S, Myers DB. 2012b. Associations between soil carbon and ecological landscape variables at escalating spatial scales in Florida, USA. *Landscape Ecol.* 27(3): 355–367.

Vasques GM, Grunwald S, Sickman JO. 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma* 146(1-2): 14–25.

Vasques GM, Grunwald S, Sickman JO. 2009. Modeling of soil organic carbon fractions using visible-near-infrared spectroscopy. *Soil Sci. Soc. Am. J.* 73(1): 176–184.

Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131(1–2): 59–75.

Webster R, Oliver MA. 2007. *Geostatistics for Environmental Scientists*. John Wiley & Sons, Chichester, England.

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(6): 823–843.

Williams P. 1987. Variables affecting near-infrared reflectance spectroscopic analysis. p. 143–167. In *Near-infrared technology in the agricultural and food industries*. American Association of Cereal Chemists, St. Paul, Minnesota.

Wu J, Jones KB, Li H, Loucks OL (Eds). 2006. *Scaling and Uncertainty Analysis in Ecology: Methods and Applications*. Springer, Dordrecht, The Netherlands.

Xiong X. 2013. *Geo-Spatial Modeling of Soil Organic Carbon and Its Uncertainty*. Ph.D. dissertation, University of Florida, Gainesville, FL.

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.

532 Table 1. Characteristics of the five fields.

Variables		Study Areas				
		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	Elevation (m)†	42.8	43.5	23.9	28.8	8.7
	Slope (%)†	1.2	1.2	2.9	2.4	0.2
Climate	Max temperature (°C)‡	27.5	27.1	26.3	27.2	29.2
	Min temperature (°C)‡	14.0	13.8	12.9	13.6	16.3
	Precipitation (mm)‡	1325	1345	1634	1360	1464
Parent Material	Surficial geology§	Cypresshead	Coosawhatchie	Citronelle	Coosawhatchie	Shelly sediments of plio-pleistocene
	Land use¶	Xeric upland forest	Mesic upland forest	Pineland	Improved pasture	Rangeland
Organism	NPP (kg C m^{-2})#	7.91	13.60	9.07	7.50	8.13
	NDVI#	3.81	7.90	3.81	9.50	4.31
	Dry biomass (kg m^{-2})††	2.76	12.50	5.53	-	6.68
Soil	Soil suborder‡‡	Psamments	Aquults-Psamments-Udepts-Udults	Psamments	Udults	Aquods
	AWC (cm cm^{-1})‡‡	1.2	2.1	1.5	2.2	1.7
	Clay content (%)‡‡	1.2	5.2	3.7	4.6	1.9
	Sand content (%)‡‡	98.6	93.2	93.1	90.8	96.8

533 Variable descriptions, abbreviations and sources: † National Elevation Dataset (NED), United States Geological
534 Survey (USGS),1999; ‡ long term maximum and minimum annual average temperature, long term annual average
535 precipitation between 1971-2000 from Parameter-elevation Regressions on Independent Slopes Model (PRISM)
536 climate group; § USGS, 1998; ¶ Florida Fish and Wildlife Conservation Commission, 2003; # net primary
537 productivity (NPP), normalized difference vegetation index (NDVI) from Moderate-Resolution Imaging
538 Spectroradiometer (MODIS) for North American Carbon Project, 2005; †† National Biomass and Carbon Dataset
539 (NBCD), 2000; ‡‡ soil suborder, available water holding capacity at 0-25cm (AWC), clay content and sand content
540 at 0-20cm from Soil Survey Geographic Database (SSURGO), Natural Resources Conservation Service (NRCS),
541 2009.

Table 2. Descriptive statistics of measured soil total carbon (original values: TC, logarithm-transformed values: logTC).

		TC (%)						logTC (log %)			
Datasets	n	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	0.030	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

CAL = the data set used to calibrate the models; VAL = the data set used to validate the models; SUB-W = the 112 observations randomly chosen from the five fields (Fig. 1); n = number of observations; SD = standard deviation, CV = coefficient of variation, skew. = skewness.

Table 3. Summary statistics for the spectral models of logTC produced by Partial Least Square Regression (PLSR), Support Vector Machine (SVM), and Random Forests (RF) derived from calibration (CAL) using 70% of all the samples (n = 392) and validation using 30% of the samples (n = 168).

	LOO Cross-Validation using CAL		Validation using VAL				
	R^2	RMSE (log %)	R^2	Bias (log %)	RMSE (log %)	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

LOO cross-validation = leave-one-out cross-validation; R^2 = coefficient of determination; RMSE = root mean squared deviations; RPD = residual prediction deviation; RPIQ = ratio of prediction error to inter-quartile range.

Table 4. Summary statistics of leave-one-out cross-validation for Partial Least Square Regression (PLSR) and Support Vector Machine (SVM) models of logTC (log %) developed in SUB-W and the five field datasets.

	PLSR		SVM	
Models	R^2	RMSE (log %)	R^2	RMSE (log %)
Model SUB-W	0.82	0.10	0.84	0.10
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

R^2 = coefficient of determination; RMSE = root mean squared deviations; RPD = residual prediction deviation; RPIQ = ratio of prediction error to inter-quartile range; SUB-W = the 112 observations randomly chosen from the five fields (Fig. 1).

Table 5. The transferability of Partial Least Square Regression (PLSR) models developed in one of the five study fields to predict the soil logTC (log %) of the other four fields.

Models	Validation datasets (n = 112)	R^2	Bias (log %)	RMSE (log %)	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

R^2 = coefficient of determination; RMSE = root mean squared deviations; RPD = residual prediction deviation; RPIQ = ratio of prediction error to inter-quartile range.

Table 6. The transferability of Support Vector Machine (SVM) models predicting soil logTC (log %) developed in one of the five study fields to predict the soil logTC (log %) of the other four fields.

Model	Test datasets (n = 112)	R^2	Bias (log %)	RMSE (log %)	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

R^2 = coefficient of determination; RMSE = root mean squared deviations; RPD = residual prediction deviation; RPIQ = ratio of prediction error to inter-quartile range.

Table 7. The down-scaling performance of the Partial Least Square Regression (PLSR) model predicting soil logTC (log %) developed at regional scale (SUB-W) predicting samples at field scales.

Model	Validation datasets (n = 112)	R^2	Bias (log %)	RMSE (log %)	RPD	RPIQ
Model SUB-W (n = 112)	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

R^2 = coefficient of determination; RMSE = root mean squared deviations; RPD = residual prediction deviation; RPIQ = ratio of prediction error to inter-quartile range; SUB-W = the 112 observations randomly chosen from the five fields (Fig. 1).

Table 8. The down-scaling performance of the Support Vector Machine (SVM) model predicting soil logTC (log %) developed at regional scale (SUB-W) predicting samples at field scales.

Model	Validation datasets (n = 112)	R^2	Bias (log %)	RMSE (log %)	RPD	RPIQ
Model SUB-W (n = 112)	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

R^2 = coefficient of determination; RMSE = root mean of the squared deviations; RPD = residual prediction deviation; RPIQ = ratio of prediction error to inter-quartile range; SUB-W = the 112 observations randomly chosen from the five fields (Fig. 1).

Table 9. The up-scaling performance of the Partial Least Square Regression (PLSR) models predicting soil logTC (log %) developed at field scale predicting samples at regional scale (SUB-W).

Models	Validation dataset	R^2	Bias (log %)	RMSE (log %)	RPD	RPIQ
Model 1	SUB-W (n = 112)	0.53	-0.12	0.22	1.09	1.97
Model 2		0.64	-0.03	0.15	1.58	2.87
Model 3		<0.01	-0.17	0.50	0.48	0.86
Model 4		0.57	0.18	0.23	1.02	1.84
Model 5		0.36	0.06	0.23	1.02	1.86

R^2 = coefficient of determination; RMSE = root mean squared deviations; RPD = residual prediction deviation; RPIQ = ratio of prediction error to inter-quartile range; SUB-W = the 112 observations randomly chosen from the five fields (Fig. 1).

Table 10. The up-scaling performance of the Support Vector Machine (SVM) models in predicting soil logTC (log %) developed at field scale predicting samples at regional scale (SUB-W).

Models	Validation dataset	R^2	Bias (log %)	RMSE (log %)	RPD	RPIQ
Model 1	SUB-W (n = 112)	0.08	-0.27	0.35	0.67	1.22
Model 2		0.19	0.15	0.27	0.90	1.63
Model 3		0.04	-0.18	0.29	0.81	1.48
Model 4		0.11	0.03	0.23	1.05	1.91
Model 5		0.15	0.21	0.31	0.77	1.40

R^2 = coefficient of determination; RMSE = root mean of the squared deviations; RPD = residual prediction deviation; RPIQ=ratio of prediction error to inter-quartile range; SUB-W = the 112 observations randomly chosen from the five fields (Fig. 1).

Table 11. Gower similarity coefficients of environmental factors among fields and across scales (SUB-W).

	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

SUB-W = the 112 observations randomly chosen from the five fields (Fig. 1).

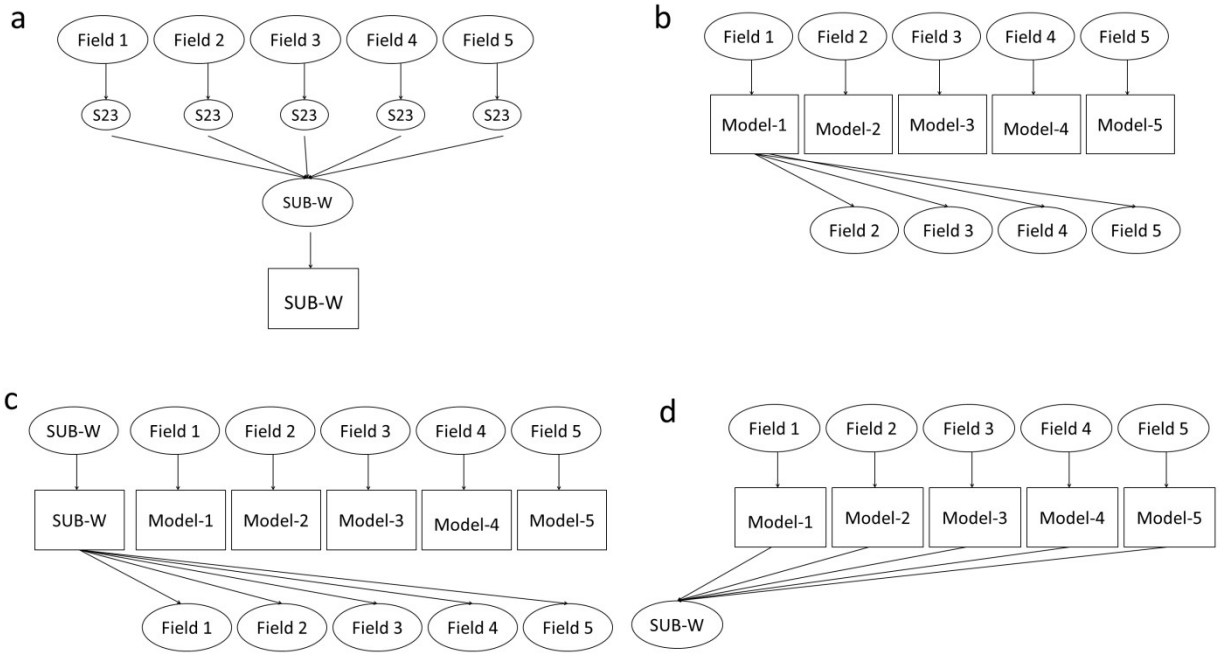

Figure 1. The principle scheme of the transferability and scaling analysis: a) the sample source of SUB-W dataset; b) transferability at field scale; c) down-scaling analysis; d) up-scaling analysis. Note: S23 in Fig.1-a represented the 23 samples randomly chosen from each of the five fields to calibrate the regional model.

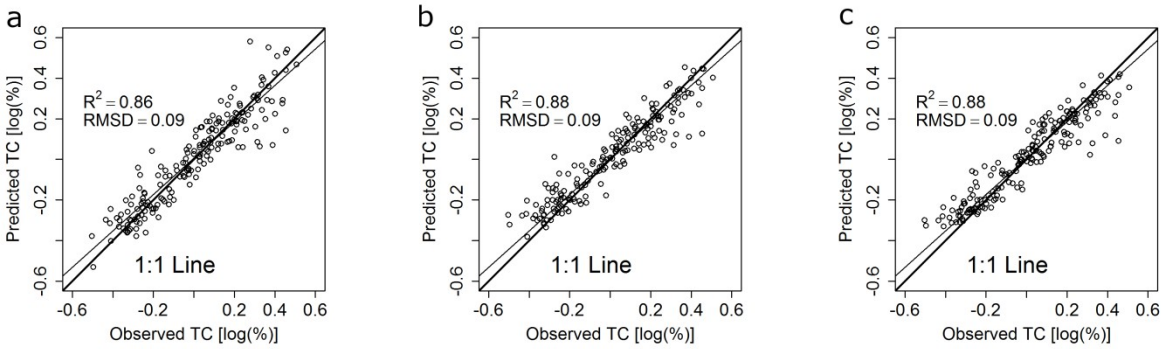

Figure 2. Predicted vs. observed logarithm transformed soil total carbon (logTC) of VAL dataset derived from: a) Partial Least Square Regression (PLSR), b) Support Vector Machine (SVM), and c) Random Forest (RF). R^2 = relation of the coefficient; RMSE = root mean of the squared deviations.

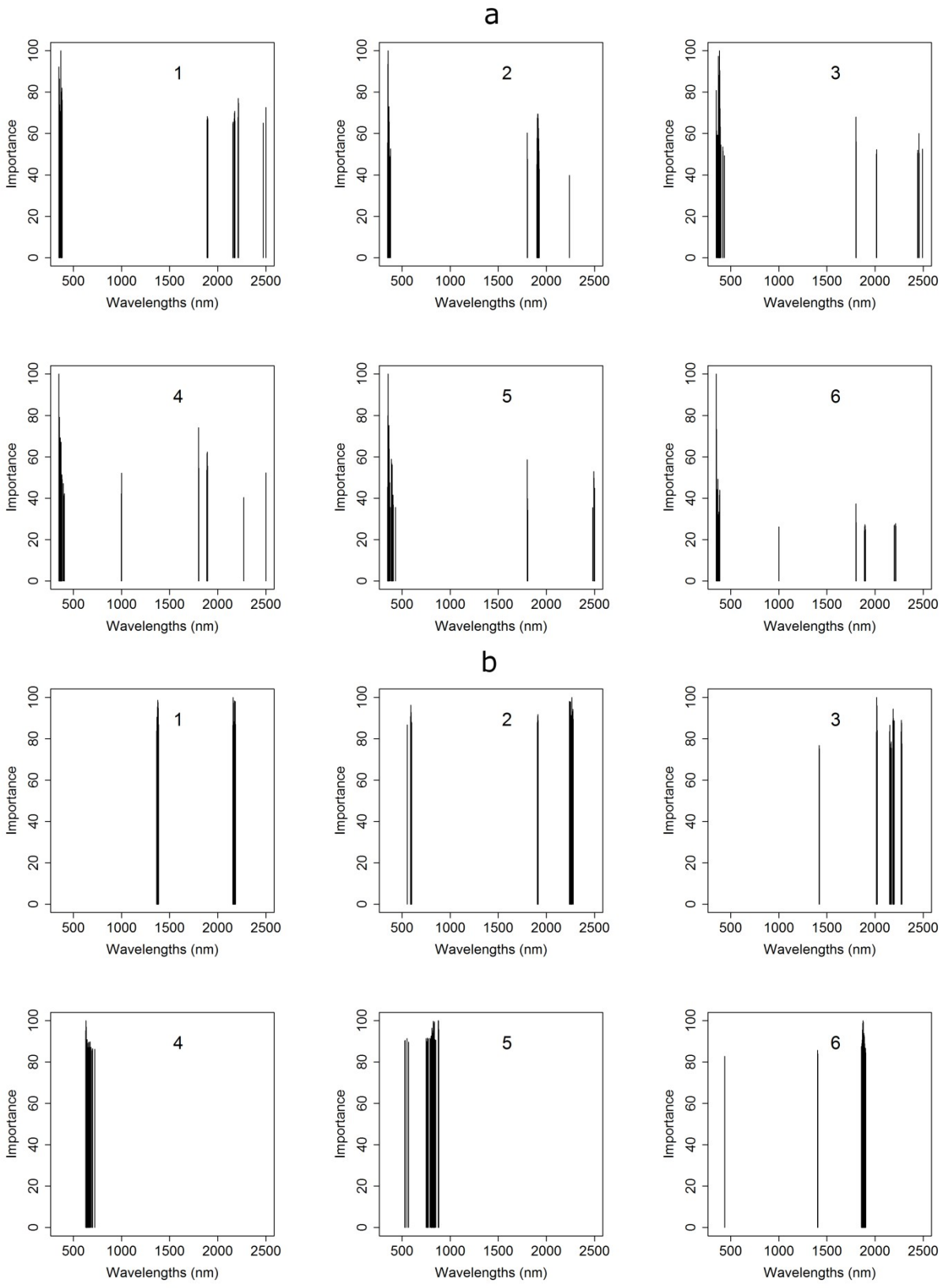

Figure 3. The importance values of top 50 predictors (wavelengths) derived from the models with a) Partial Least Square Regression (PLSR) and b) Support Vector Machine (SVM) methods: 1 is developed from the pooled SUB-W dataset; 2, 3, 4, 5, and 6 are developed from datasets of Field 1, 2, 3, 4, and 5 respectively.

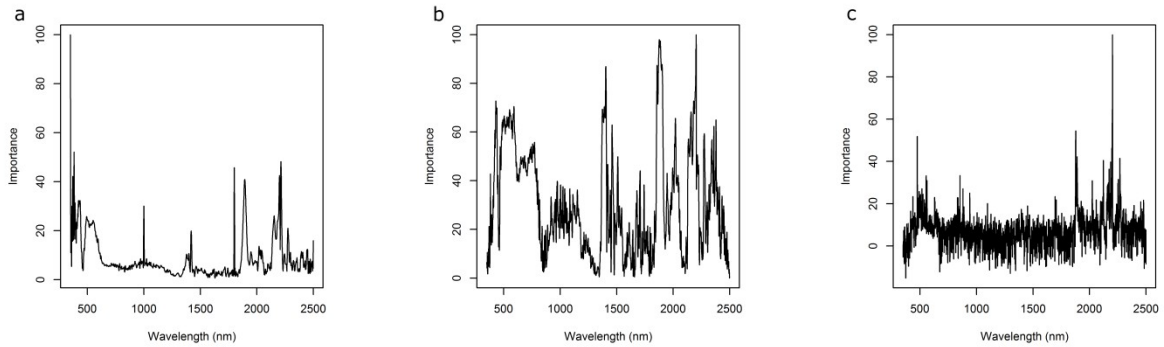

Figure 4. Variable importance values of a) Partial Least Square Regression (PLSR), b) Support Vector Machine (SVM) and c) Random Forests (RF) models derived from CAL dataset.

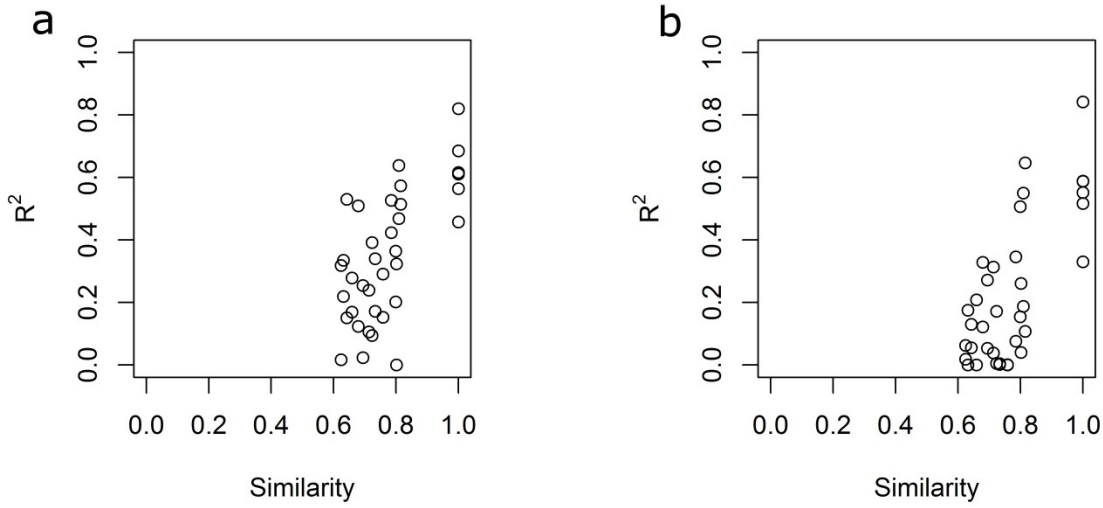

Figure 5. The coefficient of determination (R^2) of each model transferred to other fields and scale vs. the Gower similarity coefficient between the model development field/scale and the model application field/scale: a) Partial Least Square Regression (PLSR); b) Support Vector Machine (SVM).

Figure captions:

Figure 1. The principle scheme of the transferability and scaling analysis: a) the sample source of SUB-W dataset; b) transferability at field scale; c) down-scaling analysis; d) up-scaling analysis. Note: S23 in Fig.1-a represents the 23 samples randomly chosen from each of the five fields to calibrate the regional model.

Figure 2. Predicted vs. observed logarithm transformed soil total carbon (logTC) of VAL dataset derived from: a) Partial Least Square Regression (PLSR), b) Support Vector Machine (SVM), and c) Random Forest (RF). R^2 = relation of the coefficient; RMSE = root mean of the squared deviations.

Figure 3. The importance values of the top 50 predictors (wavelengths) derived from the models with a) Partial Least Square Regression (PLSR) and b) Support Vector Machine (SVM) methods: 1 is developed from the pooled SUB-W dataset; 2, 3, 4, 5, and 6 are developed from datasets of Field 1, 2, 3, 4, and 5 respectively.

Figure 4. Variable importance values of a) Partial Least Square Regression (PLSR), b) Support Vector Machine (SVM) and c) Random Forests (RF) models derived from CAL dataset.

Figure 5. The coefficient of determination (R^2) of each model transferred to other fields and scale vs. the Gower similarity coefficient between the model development field/scale and the model application field/scale: a) Partial Least Square Regression (PLSR); b) Support Vector Machine (SVM).