Upscaling of Dynamic Soil Organic Carbon Pools in a North-Central Florida Watershed

Gustavo M. Vasques
Sabine Grunwald*
Soil and Water Science Dep.
Univ. of Florida
2169 McCarty Hall
P.O. Box 110290
Gainesville, FL 32611

James O. Sickman
Dep. of Environmental Sciences
Univ. of California
Riverside, CA 92521

Nicholas B. Comerford
North Florida Research
and Education Center
Univ. of Florida
Quincy, FL 32351

Regional-scale assessment of soil C pools is essential to provide information for C cycling models, land management, and policy decisions, and elucidate the relative contribution of different C pools to total C (TC). We estimated TC and four soil C fractions, namely recalcitrant C (RC), hydrolyzable C (HC), hot-water-soluble C (SC), and mineralizable C (MC), at 0 to 30 cm across a 3585-km² mixed-use watershed in north-central Florida. We used lognormal block kriging (BK) and regression block kriging (RK) to upscale soil C using 102 training samples and compared the models using 39 validation samples. Regression kriging produced the most accurate models for TC and RC, whereas the labile C fractions (HC, SC, and MC) were best modeled by BK. Maps produced by BK showed similar spatial patterns due to the strong correlation between the labile C fractions and the similarity of their spatial dependence structure. Estimates of TC and RC were similar due to their high correlation and the similarity of their global trend models. Total soil C amounted to 27.40 Tg across the watershed, indicating the potential of these soils to store C. Recalcitrant C totalled 22.49 Tg (82% of TC), suggesting that a large amount of TC could be potentially stored for centuries to millennia. Our estimates of soil C and fractions within a mixed-use watershed in Florida highlight the importance of appropriately characterizing the inherent spatial dependence structure of soil C, as well as relevant regional environmental patterns (e.g., hydrology), to better explain the variability of soil C.

Abbreviations: BK, lognormal block kriging; HC, hydrolyzable organic carbon; MC, mineralizable organic carbon; RC, recalcitrant organic carbon; RK, regression block kriging; RMSE, root mean square error of validation; RT, regression tree; SC, hot-water-soluble organic carbon; SFRW, Santa Fe River watershed; SMLR, stepwise multiple linear regression; SOC, soil organic carbon; SOM, soil organic matter; TC, total carbon.

The demand for maps of soil properties across large regions has increased considerably in the last decades, reflecting a growing concern for the conservation of soil resources for the provision of commodities and ecosystem services, including food and fiber, water quality, and biodiversity. In this context, soil organic C (SOC) is a central property that relates to biological, chemical, and physical soil properties and processes; thus, regional estimates of SOC aid our understanding of the spatial patterns of soil ecosystem services.

Given ongoing anthropogenically induced climate change (Carbon Dioxide Information Analysis Center, 2008), it is essential to quantify the present SOC stocks and to understand how SOC behaves spatially across large regions. Spatial analyses are thus needed to assess the potential of soils to sequester C and mitigate climate change. Simply quantifying total SOC may not yield an adequate understanding of soil C dynamics, however, because C exists in several different soil pools that can be defined on the basis of stability and residence time. The recalcitrance or lability of C forms in soils depends on several environmental determinants, such as climate, topography, hydrology, land use, and other properties (Jenny, 1941; McBratney et al., 2003). Therefore, to understand the dynamics of SOC across large landscapes, multiple C pools must be considered along with their interaction with environmental landscape properties.
Many techniques have been applied to fractionate SOC or soil organic matter (SOM) into more or less stable C forms with the objective of mapping the dynamics of SOC as it interacts with local (e.g., plants and microorganisms) and global (i.e., the lithosphere, hydrosphere, and atmosphere) components (e.g., Parrot et al., 1983; Coleman and Jenkinson, 1996; Zimmermann et al., 2007). The most common fractionation techniques utilize either physical methods, chemical methods, or a combination of both (von Lützow et al., 2007). Physical fractionation usually involves separation based on particle density, aggregation or the energy of aggregation, or size (e.g., Sohi et al., 2001; Six et al., 2002; Echeverría et al., 2004; Sarkhot et al., 2007a,b), whereas chemical fractionation is commonly done by acidolysis, hydrolysis, or oxidation (e.g., Leavitt et al., 1996; Ghani et al., 2003; Silveira et al., 2008). Silveira et al. (2008) extracted labile pools of SOC using acid hydrolysis, and explained 99% of the variability of 6 mol L−1 HCl hydrolyzable SOC (HC) using microbial C biomass and SC, which demonstrated the association of the HC pool with indicators of SOC lability (i.e., availability to plants and microorganisms) in the soil. In their 18 soil samples, HC comprised 18 to 32% of the TC, whereas SC comprised 1 to 4% of the TC. Ghani et al. (2003) correlated different soil C and N fractions with land use, grazing intensity, and N and P fertilization, and found that SC was more sensitive than total SOC to land use and management practices. Moreover, SC was highly correlated with soil total carbohydrates ($R^2 = 0.88$), mineralizable N after 7 d of incubation ($R^2 = 0.86$), and soil microbial C ($R^2 = 0.84$) and N ($R^2 = 0.72$) biomass.

Incubation to quantify mineralization and immobilization rates is another method to assess the stability and lability of SOC. Mineralization rates and mean residence times (MRTs) of SOC have been associated with the chemical and physical characteristics of SOM (e.g., Franzluebbers, 1999; Alvarez and Alvarez, 2000; Paul et al., 2006; Ahn et al., 2009) and related to soil bio-physicochemical properties and processes (e.g., Alvarez et al., 1995; Causarano et al., 2008). Paul et al. (2006) used 6 mol L−1 HCl hydrolysis in 1100 samples to separate the non-hydrolyzable or recalcitrant C pool, determining that RC represented 30 to 80% of total SOC, depending on the soil type, texture, depth, and management. In addition, they measured the active and slow soil C pools after incubation, showing that active C represented 2 to 8% of SOC, with a MRT of days to months, whereas slow C comprised 45 to 65% of SOC, with a MRT of 10 to 80 yr. Causarano et al. (2008) identified differences in SOC among areas of pasture, conservation tillage, and conventional tillage. Management explained 42% of the variability of SOC, and other significant explanatory variables included clay content (5%) and mean annual temperature (1%). Total SOC was significantly correlated with all measured SOC physical and chemical fractions, including particulate organic C, microbial biomass C, and mineralizable C after 24 d of incubation.

Many soil properties have been modeled and upscaled regionally (McBratney et al., 2003). Soil C has been suggested as a good indicator of soil quality (Seybold et al., 1996) and potential soil ecosystem services. Examples of soil C studies at the regional scale include Homann et al. (1998), McKenzie and Ryan (1999), Krogh et al. (2003), Hengl et al. (2004), Minasny et al. (2006), and Vasques et al. (2010). The Soil Survey Geographic (SSURGO; Soil Data Mart) database (NRCS, 2006) provides SOM data covering most of the U.S. territory. Regional environmental determinants of TC have also been identified, including climate (e.g., Guo et al., 2006b), land use (e.g., Leifeld et al., 2005), topography (e.g., Zhong and Xu, 2009), and geology (e.g., Vasques et al., 2010). Upscaling of individual SOC chemical fractions has been rarely done, however, due to the analytical and computational costs and labor required to implement such studies. To overcome this research gap, we conducted an investigation of the spatial distribution of TC and four dynamic SOC pools within the Santa Fe River watershed (SFRW). The great variability of SOC and distinct environmental conditions in this region of Florida offer a good framework to characterize the relationship between soils and their respective soil-forming landscape factors. Our study had two major objectives: (i) to estimate the spatial patterns of TC and four SOC fractions, namely RC, HC, SC, and MC, across a large subtropical watershed using hybrid geostatistical upscaling methods; and (ii) to validate the upscaled models using independent validation data sets. We hypothesized that spatial patterns of TC and SOC fractions are a function of collocated environmental landscape properties, and thus can be spatially explained and estimated using hybrid geostatistical environmental correlation models.

**MATERIALS AND METHODS**

**Study Area**

The study was conducted in the SFRW, a 3585-km² mixed-use watershed located in north-central Florida between 29.63 and 30.21° N and 82.88 and 82.01° W. A complete description of the SFRW was presented in Vasques et al. (2008). Briefly, the most frequent soil series in the SFRW are Sapelo, Blanton, Ocala, Mascotte, and Foxworth, and the most frequent soil orders include Ultisols (47%), Spodosols (27%), and Entisols (17%; NRCS, 2009). The major land uses include pine-land (30%), wetland (14%), improved pasture (13%), upland forest (13%), and rangeland (13%; Florida Fish and Wildlife Conservation Commission, 2003).

**Field Sampling and Laboratory Methods**

A total of 141 soil samples (Fig. 1) were collected between September 2003 and January 2005 at a fixed depth of 0 to 30 cm, according to a random design stratified proportionally by the areas of soil order–land use combinations and analyzed for TC, RC, SC, and MC. Local variability was accounted for by composite sampling with four subsamples collected within a 2-m radius at each site, followed by complete homogenization before analysis. Soil organic C represents >98% of the TC in Florida (Guo et al., 2006a). Moreover, all the soils in this study had a pH below 6.0. Therefore, the soil samples were not pretreated with acid to remove carbonates.

Total C was determined by high-temperature combustion on a FlashEA 1112 Elemental Analyzer (Thermo Electron Corp., Waltham,
MA). Recalcitrant C was also measured by high-temperature combustion after samples were refluxed with 6 mol L\(^{-1}\) HCl for 16 h, according to Paul et al. (2001) and McLauchlan and Hobbie (2004). Hot-water-soluble C was measured on a Shimadzu TOC 5050 Analyzer (Shimadzu Scientific Instruments Inc., Columbia, MD) after extraction using hot water (Sparling et al., 1998; Gregorich et al., 2003) and filtration through a 0.2-µm membrane. Lastly, MC was measured during a period of 14 d of steady soil respiration inside an incubation chamber using a CO\(_2\) coulometer (UIC Inc., Joliet, IL). Hydrolyzable C was calculated by the difference between TC and RC. We confirmed that laboratory measurements were accurate by repeating the measurements of about 10% of the samples and also comparing measurements against analytical standards. The detailed sampling design and laboratory procedures were described in Vasques et al. (2008, 2009).

The hybrid geostatistical methods used in this study (compare below) either required or benefited from an approximate normal frequency distribution of the target variable. Because TC and all the SOC fractions had positively skewed lognormal distributions, they were log\(_{10}\) transformed before modeling.

**Upscaling Methods**

The laboratory SOC measurements in concentration units (mg kg\(^{-1}\)) were converted to areal units (kg m\(^{-2}\)) by multiplying the C concentration by the soil bulk density, soil thickness (30 cm), and a unit conversion factor. Rock fragments were not observed at 0 to 30 cm. The soil bulk density was estimated using a class pedotransfer function linking historical bulk density measurements (Grunwald et al., 2008) from 265 soil profiles around and within the SFRW to the corresponding soil series of the sampled soils. The observed soil bulk density values were averaged by soil series within the 0- to 30-cm depth and then attributed to the corresponding soil series encountered at the sampling sites in the SFRW. The average bulk density value across the 45 soil series included in the pedotransfer function was 1.44 ± 0.02 (mean ± standard error).

The 141 samples were randomly split into a training (102 observations) and a validation set (39 observations). Using the training samples, three upscaling approaches were compared to estimate TC and SOC fractions in the SFRW, namely BK, RK using stepwise multiple linear regression (SMLR) to model the global spatial trend, and RK using a regression tree (RT; Breiman et al., 1984) to model the global trend. In BK, ordinary block kriging was performed on the log-transformed SOC pools and the output maps converted back to the original units according to Webster and Oliver (2007). The block estimate was derived by averaging 25 estimates distributed as a 5 by 5 grid within the block with 30- by 30-m dimension. In RK, the global trend was modeled by SMLR or RT and the residuals kriged using ordinary block kriging.

\[
\gamma(b) = \begin{cases} 
    c_0 + c \left[ 1.5 \left( \frac{b}{a} \right) - 0.5 \left( \frac{b}{a} \right)^3 \right] & \text{if } b \leq a \\
    c_0 + c & \text{if } b > a \\
    0 & \text{if } b = 0
\end{cases}
\]  

where \(\gamma(b)\) is the semivariance at lag distance \(b\), \(c_0\) is the nugget effect, \(c\) is the partial sill, \(b\) is the lag distance, \(a\) is the range, and \(r\) is the effective range, where \(\gamma(b)\) achieves 95% of the total sill \((c_0 + c)\) at about 3\(a\).

The spherical (Eq. [1]) and exponential (Eq. [2]) models were manually fitted and visually compared to best approximate the experimental semivariograms observed for the different SOC pools. Upscaled output maps were created using a spatial resolution of 30 m.

The global spatial trend was modeled based on diverse environmental landscape variables covering all soil-forming factors included in the CLORPT (Jenny, 1941) and SCORPAN (McBratney et al., 2003) models, which were obtained from various sources and used as or converted to geographic information system layers. The independent variables were: 10 soil properties (taxonomic order, drainage class, hydric rating, hydrologic group, available water capacity, depth to water table, saturated hydraulic conductivity, and sand, silt, and clay contents) derived at 0 to 100 cm based on the dominant component within the map unit; 19 variables derived from Landsat 7 Enhanced Thematic Mapper Plus images (six bands, six principal components, three tasseled cap indices, and four vegetation indices); five topographic variables (elevation, slope, aspect, catchment area, and compound topographic index); mean annual temperature and precipitation; land use or land cover; physiographic region; hydrogeologic group, and aquifer vulnerability index; mean annual temperature and precipitation; land use or land cover; physiographic region; ecological region; and population count (detailed data sources were listed in Vasques et al., 2010). Stepwise multiple linear regression used
Table 1. Descriptive statistics of measured soil organic C pools: total organic C (TC), recalcitrant organic C (RC), hydrolyzable organic C (HC), hot-water-soluble organic C (SC), and mineralizable organic C (MC).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>TC</th>
<th>RC</th>
<th>HC</th>
<th>SC</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean, kg m⁻²</td>
<td>6.26</td>
<td>4.67</td>
<td>1.59</td>
<td>0.34</td>
<td>0.05</td>
</tr>
<tr>
<td>Median, kg m⁻²</td>
<td>4.57</td>
<td>3.30</td>
<td>1.28</td>
<td>0.29</td>
<td>0.04</td>
</tr>
<tr>
<td>SD, kg m⁻²</td>
<td>8.04</td>
<td>6.91</td>
<td>1.33</td>
<td>0.27</td>
<td>0.04</td>
</tr>
<tr>
<td>CV, %</td>
<td>128.46</td>
<td>148.12</td>
<td>83.86</td>
<td>79.33</td>
<td>80.85</td>
</tr>
<tr>
<td>Skewness</td>
<td>5.43</td>
<td>5.47</td>
<td>4.74</td>
<td>5.38</td>
<td>3.53</td>
</tr>
</tbody>
</table>

an F probability of 0.05 to include or exclude variables from the model and was conducted in SPSS (SPSS Inc., Chicago, IL). The RT models used 10-fold cross-validation on the training samples and were derived in CART (Salford Systems, San Diego, CA). Block kriging was conducted in ISATIS (Geovariances, Avon, France).

Accuracy assessment of the upscaled SOC pools was conducted by comparing the estimated values in the final output maps produced by BK, RK/SMLR, and RK/RT with the independent validation samples. The best model of each SOC pool was identified based on the root mean square error of independent validation (RMSEv):

\[
RMSE_v = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad [3]
\]

and residual prediction deviation (RPD; Williams, 1987):

\[
RPD = \frac{SD_{val}}{\sqrt{RMSE_v(n/(n-1))}} \quad [4]
\]

where \(\hat{y}_i\) are the predicted values, \(y_i\) are the observed values, \(n\) is the number of predicted or observed values in the validation set with \(i = 1, 2, \ldots, n\), and \(SD_{val}\) is the standard deviation of the validation set.

RESULTS AND DISCUSSION

Descriptive Statistics

Total C (~total SOC) varied from 1.17 to 63.88 kg m⁻², with a mean of 6.26 kg m⁻² and a median of 4.57 kg m⁻² (Table 1). After logarithmic transformation, the mean and median values resembled each other and the original skewness of 5.43 was reduced to 1.19, approximating a normal distribution. The range of the TC training sample (102 observations) encompassed the range of the validation sample (39 observations), which is advisable to avoid extrapolation. Their frequency distribution was dissimilar, however, as TC values in the training sample varied more and had a more skewed histogram (not shown).

Similar trends were observed on the measured SOC fractions. The range of the training sample encompassed that of the validation sample for all SOC fractions but HC, and the frequency distribution of the training sample was more positively skewed and more variable than that of the validation sample for all SOC fractions. Logarithmic transformation reduced the variability (i.e., the coefficient of variation) of all SOC fractions except HC.

Correlation coefficients among TC and the SOC fractions were reported in concentration units (mg kg⁻¹) in Vasques et al. (2009) and were all significant at the \(P = 0.01\) level, varying from 0.53 between HC and MC to 0.98 between TC and RC. Similarly, in the present study, all correlations among TC and SOC fractions expressed in content units (kg m⁻²) were significant at the 0.01 confidence level. Correlations expressed in concentration units (Vasques et al., 2009) are not influenced by bulk density variations. Vasques et al. (2009) noted that the concentration of SOC fractions in the soil was inversely related to their lability. Thus, the concentration (and recalcitrance) of SOC fractions decreased in the following order: RC > HC > SC > MC. Recalcitrant organic C comprised about 75% of the TC, while HC comprised the remaining 25%. Together, the most labile fractions, SC and MC, comprised no more than 5% of the TC.

Upscaling of Soil Organic Carbon Pools

Among the five SOC pools analyzed, only TC and RC were best modeled by RK, according to the RMSEv and RPD (Table 2). The three labile C fractions (HC, SC, and MC) were best modeled by BK. Based on the best upscaled maps, the average TC stock at 0 to 30 cm in the SFRW was 7.64 kg m⁻² (Table 2) and the total stock across the watershed amounted to 27.40 Tg (Fig. 2a), of which 22.49 Tg consisted of RC (Fig. 2b), which has a relatively long residence time in most soils. In contrast, upscaled labile C fraction stocks were, on average, 2.59, 0.55, and 0.08 kg m⁻² for HC, SC, and MC, respectively (Fig. 2c, 2d, and 2e).

The SMLR and RT global trend models (Table 3) explained 43 and 69%, respectively, of the variability of TC in training mode, but no more than 18% in validation mode. Interestingly, not a single predictor had a strong correlation with TC or any of the SOC fractions (correlations not reported). All the assumptions of SMLR were met, except for some collinearity among the predictors and minor heteroscedasticity in the residuals of some of the models. Correlation between the residuals and the dependent variables due to the remaining unexplained variability in the dependent variables was minimized by kriging the residuals.

Maps produced for TC (Fig. 2a) and RC (Fig. 2b) showed similar patterns. Recalcitrant C constitutes the major portion of TC, explaining the high correlation (0.98) between these two SOC pools. As a consequence, the spatial distribution of RC was influenced by the same environmental factors that control the distribution of TC. In effect, their SMLR global trend models selected the same predictors with similar regression coefficients (Table 3), and their spatial dependence structures were also very similar (Table 4).

High TC and RC occurred in wetland areas around lakes and depressions, and along the Santa Fe River (dark spots in Fig. 2a and 2b). Soils that are saturated with water for long periods of time have slower decomposition rates of litter and SOM compared with dry soils (Reddy and Patrick, 1975; Bouchard and Cochran, 2006; Qualls and Richardson, 2008) where aerobic microbial communities use more efficient metabolic pathways (Kadlec and Knight, 1996). Wetlands are also expected to have higher net primary productivity than adjacent upland ecosystems.
Areas of low net productivity associated with excessively drained quartzarenic Entisols were located in the southwest portion of the watershed on top of karst terrain. A clear change from medium TC and RC stocks to low stocks in the east–west direction matches an abrupt change in elevation and parent material marked by an escarpment (the Cody Scarp) that cuts across the watershed in the north–south direction (Fig. 1). Medium TC and RC stocks occurring in the uplands east of the escarpment are associated with more fertile soils (mainly Ultisols and Spodosols) and terrain of slower infiltration compared with the karst terrain in the western lowlands that have generated Entisols with high infiltration rates that promote lower natural productivity and faster loss of dissolved organic C. An exception is the southwesternmost portion of the watershed, which contains pronounced wetlands and thus hydric soils (Aquods, Aquults, Aqualfs, and Sapristis) that are high in TC.

For comparison, we estimated SOC stocks in the SFRW using SSURGO data (NRCS, 2009) and the Soil Data Viewer application (NRCS, 2007). We used the option to calculate the weighted averages of bulk density and SOM from 0 to 30 cm using the component percentage within the map unit as the weight. The total SOC stock across the SFRW estimated using SSURGO was 27.08 Tg, which was very close to our TC estimate of 27.40 Tg by RK/SMLR. The average SSURGO SOC stock was 7.56 kg m\(^{-2}\), compared with our average TC estimate of 7.64 kg m\(^{-2}\).

The spatial patterns of the labile C fractions (HC, SC, and MC) were very similar (Fig. 2c, 2d, and 2e) but differed from those of TC and RC because of the different upsampling method (BK). Similar patterns were a result of the correlation among those of TC and RC because of the different upsampling method (MC). Similar patterns were a result of the correlation among those of TC and RC because of the different upsampling method (MC). The spatial dependence structure identified by the BK model for TC was very similar to that of TC and RC, as expected (Table 4). From a pedologic perspective, these similar spatial patterns result from the fact that the three labile C fractions behave in accordance with the same environmental processes and stressors, e.g., temperature, microbial activity, hydrology, etc. In other words, if HC is high due to a set of favorable landscape conditions, then probably SC and MC will benefit from the same conditions and likewise show high values.

In spite of the differences between the labile C maps and the TC and RC maps, TC and RC were significantly correlated with all labile C fractions, suggesting some commonalities in terms of spatial trends. For example, the same hot spots of high C present in the BK maps were also evident in the TC and RC maps, corresponding to wetland areas of Histosols in the east to central south and Spodosols in the west to central north, respectively. Regions of medium C stock, mainly on Ultisols, were widespread in the central portion of the SFRW in all maps. Furthermore, regions of low C were present in the southwest, associated with low-productivity areas dominated by excessively drained Entisols, and in the easternmost portion of the watershed, where Psammists have formed.

One disadvantage of BK relative to RK was the introduction of undesirable map artifacts. One of the artifacts is the appearance of linear features caused by the search neighborhood optimized to fit the spatial autocorrelation range that included a limited number of neighbors (10 observations). On the other hand, extending the search neighborhood way beyond the spatial autocorrelation range would lead to artificial smoothing due to the incorporation of observations that are spatially unrelated to the site being estimated. A second map artifact is the appearance of round features around sampling sites. This is due to the very nature of BK, which is an exact linear interpolator and, as such, can give too much weight to observations closer to the unsampled location relative to observations farther away. The stronger the spatial dependence of the SOC property (compare Table 4) the more prominent are the round features. Nevertheless, even if kriging artifacts were present, the spatial variability of SOC at the regional scale was to some extent properly characterized by BK based on the smallest estimation errors. Two examples are the correct placement of hot spots of high SOC around wetlands, which indeed store high amounts of SOC, and the correct depiction of low-C areas associated with karst terrain in the southwest.

The spatial dependence structure identified by the BK models of TC and the SOC fractions was more important in most of the cases (three out of five properties) than their relationship with other environmental factors. Given the moderate to low explanatory power of the global trend models (Table 3), the superiority of BK over RK for most of the SOC pools was not surprising and resulted in similar distribution patterns of the three labile C fractions (HC, SC, and MC), whose spatial dependence structures resembled each other (Table 4).

Since HC, SC, and MC were statistically correlated (\(P < 0.05\)), it was not surprising to observe similar geostatistical behavior among them. First, the exponential model was chosen to fit the experimental semivariograms for all SOC pools. Second, the lag

<table>
<thead>
<tr>
<th>Pool</th>
<th>Avg. estimated soil organic C (\text{kg} \text{ m}^{-2})</th>
<th>RMSE(_v)†</th>
<th>RPD‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC</td>
<td>10.33 7.64§</td>
<td>8.18 6.34 3.05 3.69</td>
<td>0.38 0.79 0.65</td>
</tr>
<tr>
<td>RC</td>
<td>7.97 6.27</td>
<td>7.06 5.35 2.84 3.22</td>
<td>0.38 0.72 0.63</td>
</tr>
<tr>
<td>HC</td>
<td>2.59 3.50</td>
<td>3.48 1.27 2.00 2.07</td>
<td>0.50 0.32 0.30</td>
</tr>
<tr>
<td>SC</td>
<td>0.55 22.96</td>
<td>2.10 0.29 18.75 1.73</td>
<td>0.40 0.01 0.07</td>
</tr>
<tr>
<td>MC</td>
<td>0.08 23.62</td>
<td>1.93 0.05 19.27 1.83</td>
<td>0.48 0.00 0.01</td>
</tr>
</tbody>
</table>

† RMSE\(_v\), root mean square error calculated using the validation set.
‡ RPD, residual prediction deviation.
§ Results obtained from the best models are shown in bold type.
sizes chosen to approximate the experimental semivariogram were either 2150 or 2200 m, and the corresponding number of lags was 18 or 17, respectively. Third, their semivariogram ranges were on the same order of magnitude, varying from 9664 m (SC) to 12,053 m (HC). And fourth, their spatial dependence was also similar, as evidenced by the small nugget/sill ratios, ranging from about 13% (SC) to about 19% (MC), indicating strong spatial dependence (Cambardella et al., 1994). The only exception was HC, whose nugget/sill ratio was 39%, indicating a moderately strong spatial autocorrelation. (Table 4).
Table 3. Stepwise multiple linear regression (SMLR) models and variables selected by regression tree (RT) models of the global trend of soil organic C pools (total organic C [TC], recalcitrant organic C [RC], hydrolyzable organic C [HC], hot-water-soluble organic C [SC], and mineralizable organic C [MC]), with coefficients of determination calculated using the training ($R^2$) and validation sets ($R^2_v$).

<table>
<thead>
<tr>
<th>Pool</th>
<th>SMLR Model</th>
<th>$R^2$</th>
<th>$R^2_v$</th>
<th>Selected variables $\pm$ standard error</th>
<th>$R^2$</th>
<th>$R^2_v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC</td>
<td>$0.69 \pm 0.03 + 0.41 \pm 0.06$ WETLAND $- 0.24 \pm 0.07$ DRAINEXC $- 0.12 \pm 0.06$ PLEISTOCENE</td>
<td>0.43</td>
<td>0.12</td>
<td>LULC, DEPWATTBL, EYEGEOCL, DRAINCL, GEOUNIT, SLOPE(3×3), ETMVEGIND(7×7), CTI(7×7), PPT</td>
<td>0.69</td>
<td>0.18</td>
</tr>
<tr>
<td>RC</td>
<td>$0.53 \pm 0.03 + 0.47 \pm 0.07$ WETLAND $- 0.33 \pm 0.06$ DRAINEXC $- 0.15 \pm 0.07$ PLEISTOCENE</td>
<td>0.47</td>
<td>0.11</td>
<td>DEPWATTBL</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>HC</td>
<td>$-0.65 \pm 0.26 + 0.36 \pm 0.08$ WETLAND $+ 0.003 \pm 0.001$ ETM5B(3×3) $- 0.002 \pm 0.001$ DEPWATTBL $- 0.07 \pm 0.02$ ETMPC6(3×3) $+ 0.03 \pm 0.02$ ETMPC6</td>
<td>0.31</td>
<td>0.01</td>
<td>DEPWATTBL, LULC</td>
<td>0.22</td>
<td>0.05</td>
</tr>
<tr>
<td>SC</td>
<td>$-0.50 \pm 0.05 - 0.002 \pm 0.005$ KSAT $+ 0.20 \pm 0.05$ WETLAND $+ 0.14 \pm 0.04$ HYDRIC</td>
<td>0.43</td>
<td>0.14</td>
<td>HYDGROUP</td>
<td>0.21</td>
<td>0.13</td>
</tr>
<tr>
<td>MC</td>
<td>$-2.06 \pm 0.19 - 0.002 \pm 0.001$ KSAT $+ 0.01 \pm 0.002$ ETMPC2(7×7) $+ 0.01 \pm 0.004$ CTI $- 0.14 \pm 0.05$ FOREST $+ 0.13 \pm 0.05$ HYDGEODEP</td>
<td>0.42</td>
<td>0.05</td>
<td>CLAY</td>
<td>0.18</td>
<td>0.01</td>
</tr>
</tbody>
</table>

† Regression coefficient ± standard error is indicated. WETLAND, wetland; DRAINEXC, excessively drained soil; PLEISTOCENE, geologic formation that originated during the Pleistocene; ETM5B(3×3), reflectance of Band 5 (shortwave infrared) of Landsat Enhanced Thematic Mapper Plus (ETM+); ETMPC6(3×3), Principal Component (PC) 6 derived from Landsat ETM+; ETMPC6, PC 6 derived from Landsat ETM+; KSAT, soil saturated hydraulic conductivity; HYDRIC, hydric soil; ETMPC2(7×7), PC 2 derived from Landsat ETM+ averaged within a 7- by 7-pixel window; CTI, compound topographic index; FOREST, forest; HYDGEODEP, hydrogeologic class (coastal deposits).
‡ Variogram parameters of the residuals of the stepwise multiple linear regression global spatial trend model.
§ The most important variable is shown in bold type. Variable importance was based on the standardized coefficients in SMLR and on an index of relative importance in RT (neither reported).

Table 4. Semivariogram parameters of the fitted exponential models of the soil organic C pools.

<table>
<thead>
<tr>
<th>Pool</th>
<th>Lag options</th>
<th>Nugget effect</th>
<th>Sill</th>
<th>Effective range</th>
<th>Nugget/sill</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogTC</td>
<td>3000 m</td>
<td>0.0159</td>
<td>0.0879</td>
<td>11,579</td>
<td>18.14</td>
<td>0.15</td>
</tr>
<tr>
<td>LogTC#</td>
<td>2150 m</td>
<td>0.0090</td>
<td>0.0513</td>
<td>10,041</td>
<td>17.47</td>
<td>0.14</td>
</tr>
<tr>
<td>LogRC</td>
<td>2200 m</td>
<td>0.0260</td>
<td>0.1162</td>
<td>11,700</td>
<td>22.38</td>
<td>0.21</td>
</tr>
<tr>
<td>LogRC#</td>
<td>2200 m</td>
<td>0.0179</td>
<td>0.0629</td>
<td>7219</td>
<td>28.39</td>
<td>0.17</td>
</tr>
<tr>
<td>LogHC</td>
<td>2150 m</td>
<td>0.0363</td>
<td>0.0923</td>
<td>12,053</td>
<td>39.33</td>
<td>0.01</td>
</tr>
<tr>
<td>LogSC</td>
<td>2150 m</td>
<td>0.0070</td>
<td>0.0540</td>
<td>9664</td>
<td>12.96</td>
<td>0.22</td>
</tr>
<tr>
<td>LogMC</td>
<td>2150 m</td>
<td>0.0138</td>
<td>0.0738</td>
<td>10,293</td>
<td>18.70</td>
<td>0.26</td>
</tr>
</tbody>
</table>

† LogTC, log$_{10}$ of total organic C; LogRC, log$_{10}$ of recalcitrant organic C; LogHC, log$_{10}$ of hydrolyzable organic C; LogSC, log$_{10}$ of hot-water-soluble organic C; LogMC, log$_{10}$ of mineralizable organic C.
¤ Variogram parameters of the residuals of the stepwise multiple linear regression global spatial trend model.
Zhang and McGrath (2004) found a range of 100 km mapping SOC at 0 to 10 cm in southeast Ireland (~15,460 km²), based on 220 observations (~0.0001 sample ha⁻¹). Liu et al. (2006) observed a range of 632 km for SOC at 0 to 20 cm in a 3435-km² cropland region in northeast China, using 354 observations (~0.001 sample ha⁻¹).

These prior studies confirm the trend that the larger the study area, the larger the spatial correlation distance, i.e., the range. As the study area increases, landscape patterns occurring across longer distances that influence the spatial structure of SOC are identified by the semivariogram. Furthermore, because the number of samples collected per unit area usually decreases with an increase in size of the study area, there are often not enough samples in the close range to adequately characterize the short- to medium-range variability of SOC. As a consequence, the limited number of samples spread out across the study area can only capture the long-range variability of SOC, which was probably the case for the present study with ~0.0003 sample ha⁻¹.

In the SFRW, the spatial dependence structure was a stronger determinant of the spatial distribution of all labile SOC fractions than their correlation with collocated environmental properties. In a companion study, we upscaled SOC at different depth intervals (Vasques et al., 2010) and observed that BK was also preferred instead of RK for two out of five depth intervals, suggesting that characterizing the spatial dependence structure of SOC was important to capture its spatial patterns in the SFRW. For TC and RC, our findings confirm other studies that compared variations of kriging, and other upsampling methods, to interpolate soil C or SOM. For example, López-Granados et al. (2002, 2005) obtained more accurate results using RK than ordinary kriging (OK) or simple linear regression. Mueller and Pierce (2003) preferred kriging with an external drift (KED) over SMLR, OK, co-kriging (COK), or universal kriging (UK) when using a more complete sampling grid (134 observations), but UK performed best when the sampling size was reduced to 38 samples. They used elevation as the external drift variable. Finally, Simbahan et al. (2006) compared multiple variations of kriging (OK, COK, KED, and RK) to upscale SOC and found that RK outperformed the other methods in all three fields. Kriging with an external drift was the second most accurate method, followed by COK, while OK produced the worst results in all three fields. Kriging with the close range to adequately characterize the short- to medium-range variability of SOC. As a consequence, the limited number of samples spread out across the study area can only capture the long-range variability of SOC, which was probably the case for the present study with ~0.0003 sample ha⁻¹.

Conversely, in the case of HC, SC, and MC, our study agrees with Terra et al. (2004), who obtained the lowest RMSE, for SOC using OK compared with COK, RK, and multiple linear regression. They mapped SOC in a cotton field in central Alabama and tested three sampling designs with escalating grid intervals, containing 496, 248, and 68 observations, respectively. Ordinary kriging provided the most accurate maps for all sampling designs, with RMSE values of 0.30 kg m⁻² for the sampling designs with 496 and 248 observations and 0.36 kg m⁻² for the sampling design with 68 observations.

The SFRW is comprised of many types of soils and land uses; thus, a model to upscale SOC across this highly complex mixture of ecotypes should, at least in part, account for this variability. Therefore, we expected that RK would perform better than BK, which was not the case for three out of the five SOC pools investigated. Based on these findings, in the case of TC and RC, our hypothesis that SOC is a function of collocated environmental variables is more defensible than in the case of the labile SOC fractions, whose spatial dependence structure alone was a better predictor than environmental covariates. For all the SOC pools, however, there was only weak evidence supporting the influence of the landscape factors included in this study on the spatial distribution of SOC, as the global trend models did not explain more than 18% of the SOC variation in validation mode.

Two explanations for this observation are possible. First, the variability of TC and the SOC fractions could have been inherently represented under the BK models. This would be reasonable since the upscaled maps agree on the hot spots of SOC, which were concentrated around wetlands and Spodosols. An alternate explanation for the superiority of BK over RK could be that the variability of environmental properties was so large in the SFRW that environmental patterns were either not apparent or not linked strongly enough to the distribution of the SOC pools. This was the case for hydrology, land use, and soil type (more specifically the presence of organic and spodic horizons), which seem to be the most important landscape determinants of soil C in this sandy and flat watershed that is relatively young in geologic age. Along these lines, not a single environmental factor showed a high correlation with any of the SOC pools, resulting in a lack of strong soil-environmental trends within the watershed. This could potentially be alleviated if more samples were collected to better capture the relationships between soil C and environmental variables. Moreover, the results of this study suggest that the inclusion of better hydrologic properties in the model would be beneficial. The problem is that estimating hydrologic patterns across the watershed would be more difficult and probably less accurate than estimating SOC in the first place.

From a C sequestration perspective, it is desirable to accrete RC, which is the most complex and resistant SOC fraction, lasting in the soil from decades to thousands of years (Quideau, 2006; Rice, 2006). Research has shown that RC constitutes a major proportion of the TC in many types of soils (Paul et al., 2006; Silveira et al., 2008; Vasques et al., 2009). Labile SOC fractions, on the other hand, are less stable in the environment because they are composed of simpler molecules, including carbohydrates and proteins, also constituting the soil microbial biomass (Balaria et al., 2009). Within a relatively short period of time, labile SOC fractions can be converted to more recalcitrant forms through biochemical transformations (i.e., humification), used to sustain soil microbial communities and ultimately plant growth, or simply lost by leaching. Because they are less protected than RC, they could be more sensitive to environmental change, including climate change, land use shifts, and changes in land management. Therefore, conservation of soil C will depend on fostering the conversion of labile C to RC while minimizing its loss. Similarly, RC is stable as long as it stays protected from physically, biologically, and chemically induced decomposition.
Disturbance of soils and ecosystems can disrupt the protection of RC, promoting its conversion to less stable forms and ultimately its loss to the environment.

CONCLUSIONS

The total soil C in the SFRW added up to 27.40 Tg in the upper 30 cm. Of this total, recalcitrant forms of C accounted for >75%, suggesting that the majority of C stored in the soils of this watershed could potentially remain stable for decades to thousands of years.

Independent validation of the upscaled models indicated a preference of RK for TC and RC, and BK for the three labile SOC pools (HC, SC, and MC), the latter probably a consequence of the lack of strong correlations between these pools and environmental properties in the SFRW. In spite of the moderate to low predictive capabilities of the regression models of SOC pools, however, they consistently selected environmental properties related to hydrology (e.g., wetlands, depth to the water table, and saturated hydraulic conductivity), thereby suggesting that soil C patterns in the SFRW were significantly influenced by hydrologic patterns. In spite of the low predictive power, the upscaled maps of TC and the SOC fractions depicted clear spatial trends in the SFRW, which has important implications for sustainability and environmental quality.

Some of the soil–landscape relationships in the SFRW were captured by our models; however, some of the variability was left unexplained, which will need further attention in future studies in the subtropical southeastern United States and similar landscapes to isolate the causes of soil C (de)stabilization, accretion, and depletion, aiming to promote its long-term storage. In the context of current climate change trends, it is essential to assess both the total amount of soil C storage and the amount held in SOC pools of varying residence time and stability. Because SOC fractions are more costly and laborious to measure, the TC map could offer a good indication of the main trends and most important hot spots of C because TC was strongly correlated with measured SOC fractions. In other words, TC could be used as an alternative indicator in biogeochemical studies in lieu of SOC fractions that are more costly and laborious to measure. Moreover, the high resemblance among upscaled labile SOC fraction maps imply that measuring one of the SOC fractions in the SFRW will result in some predictive information about the other ones.

Our assessment of the SFRW can provide some guidelines to estimate soil C pools in larger regions, including the state of Florida and the southeastern United States. First, the importance of the spatial dependence structure, in other words the local soil C variability, to upscale TC and SOC fractions in the SFRW was demonstrated when comparing among upscaling methods. Second, it became apparent that soil–landscape correlations based on the available environmental data were not strong enough to derive high-quality prediction maps of SOC pools, compared with other regions where landscape gradients (e.g., topography or parent material) control the variation of SOC. Finally, if correlations among TC and the measured SOC fractions are strong, the same spatial trends can be expected if the dominant driver of spatial variation is the dependence structure (i.e., spatial autocorrelation) of the SOC fraction.

ACKNOWLEDGMENTS

Funding was provided by the Cooperative Ecosystem Service Unit, NRCS. Funding was also provided by the large-scale project "Geo-temporal estimation and visualization of nitrogen and other soil properties in the mixed-use Santa Fe River watershed", Nutrient Science for Improved Watershed Management Program, USDA, and by the University of Florida Alumni Fellowship program. We would like to thank Carolyn Olsen, Jonathan Hempel, Chunhao Xu, Mi-Youn Ahn, Yu Wang, Wade Hurt, Gregory Bruland, Christine Bliss, and Sanjay Lambal for their support in this project.

REFERENCES

CDAAC, Oak Ridge, TN.
Guo, Y., R. Amundson, P. Gong, and Q. Yu. 2006a. Quantity and spatial