Assessment of Carbon Stocks in the Topsoil Using Random Forest and Remote Sensing Images

Jongsung Kim and Sabine Grunwald*

Abstract

Wetland soils are able to exhibit both consumption and production of greenhouse gases, and they play an important role in the regulation of the global carbon (C) cycle. Still, it is challenging to accurately evaluate the actual amount of C stored in wetlands. The incorporation of remote sensing data into digital soil models has great potential to assess C stocks in wetland soils. Our objectives were (i) to develop C stock prediction models utilizing remote sensing images and environmental ancillary data, (ii) to identify the prime environmental predictor variables that explain the spatial distribution of soil C, and (iii) to assess the amount of C stored in the top 20 cm soils of a prominent nutrient-enriched wetland. We collected a total of 108 soil cores at two soil depths (0–10 cm and 10–20 cm) in the Water Conservation Area 2A, FL. We developed random forest models to predict soil C stocks using field observation data, environmental ancillary data, and spectral data derived from remote sensing images, including Satellite Pour l’Observation de la Terre (spatial resolution: 10 m), Landsat Enhanced Thematic Mapper Plus (30 m), and Moderate Resolution Imaging Spectroradiometer (250 m). The random forest models showed high performance to predict C stocks, and variable importance revealed that hydrology was the major environmental factor explaining the spatial distribution of soil C stocks in Water Conservation Area 2A. Our results showed that this area stores about 4.2 Tg (4.2 Mt) of C in the top 20 cm soils.

Core Ideas

• Remote sensing-supported models produced excellent predictions in a carbon-rich system.
• Finer spatial resolution images did not produce more accurate soil carbon predictions.
• Vegetation indices (SPOT, Landsat ETM+, and MODIS) served as major predictors in soil carbon models.

Globally, soils contain about 1400 to 2300 Pg of carbon (C), and about 20 to 30% of the soil C is stored in wetlands (Mitsch and Gosselink, 2007; Lal, 2008). The role of wetlands is critical in global biogeochemical cycles, since wetland soils are able to exhibit both consumption and production of greenhouse gases, such as methane and carbon dioxide. Hence, wetlands can shift over time between acting as a C sink or a source, depending on internal C cycling, environmental conditions (e.g., hydroperiods and nutrient status), management, and imposed stressors (e.g., global climate change).

Temperature and precipitation are key factors that control the C budget in terrestrial and aquatic ecosystems (Chen et al., 2013; Gatti et al., 2014). Global mean temperature is projected to increase by 4.8°C on average by the end of the 21st century (Hiraishi et al., 2014). Especially, temperature in the southeastern United States is expected to increase by 1.5 to 5°C on average, and precipitation is projected to decrease by approximately 10 to 20% (Karl et al., 2009; Obesekera et al., 2015), which may lead to large potential losses of C from wetland soils, shifting them from a C sink to a source. Although uncertainty exists in climate projections, numerous studies have shown that climate change could substantially affect the C cycle regionally and globally (Ciais et al., 2005; Frank et al., 2010; Dinsmore et al., 2013; Reichstein et al., 2013). A good understanding of the spatial distribution of soil C stocks is necessary for assessing the effect of future climate changes (Köchy et al., 2014). The assessment of soil C stocks in wetlands are particularly important since large stocks of C become exposed to aerobic conditions when wetlands drain (e.g., due to drier climate), and this climate sensitivity cannot be modeled with data dominated by upland ecosystems (Davidson and Janssens, 2006).

Digital soil mapping and modeling (DSMM) has shown improved accuracy in assessing spatial patterns and stocks of soil C as compared with traditional soil surveys (Thompson and Kolka, 2005; Simbahan et al., 2006; Vasques et al., 2010; Wiesmeier et al., 2011). Moreover, remote sensing (RS)-informed
soil prediction models have been successful in improving the predictive power in DSMM (Minsay and Hartemink, 2011; Mulder et al., 2011) because of their capabilities to directly or indirectly infer biotic properties. Hence, RS-informed DSMM has much value in wetlands, where field sampling is challenging and site access is limited.

Wetlands cover approximately 50% of Florida in the United States, and the Greater Everglades is the largest wetland among them. Studies of the Everglades have focused primarily on restoration assessment (Noe et al., 2001; Bernhardt and Willard, 2009). However, spatially-explicit soil C stocks have been less studied, despite the important role that soil C plays in the C cycle. Therefore, our research focused on a subsystem of the Everglades to address the following objectives: (i) to develop accurate C stock prediction models utilizing soil organic C measurements, RS images, and environmental ancillary data; (ii) to identify the most influential environmental predictor variables that explain the spatial distribution of soil C; (iii) to assess the amount of C stored in the top 20-cm soils of a prominent, nutrient-enriched wetland; and (iv) to discuss its role as a C sink or a C source.

Materials and Methods

Study Area

The Everglades in Florida (8211 km²) was fragmented into several hydrologic units, including Water Conservation Areas (WCAs) and the Everglades National Park (ENP) in the 1960s. This study was conducted in WCA-2A, the smallest hydrologic unit, which covers about 418 km² and is located in the northern part of the Everglades. In this region, the climate is characterized as subtropical, with an average annual precipitation of 1297 mm (USGS, 2009) and an average annual temperature of 23°C (NOAA, 2009). Soils and vegetation patterns are influenced by wet and dry periods, nutrient influx, fire, and other environmental stresses (Rivero et al., 2009). The majority of the soils in WCA-2A are Haplosapristis (Histosols dominated by Sapristis), which are characterized by highly decomposed organic soil materials. The average depth of soil in WCA-2A is 15 cm, and the average depth of the O horizon (consisting mainly of organic material) is 136.4 cm (Kim et al., 2012). The Everglades peat and Loxahatchee peat formations are common types of peat in the region. Soils in the Everglades peat are generally brown to black with minimal content of minerals, while soils in the Loxahatchee peat are lighter colored with remains of the roots and rhizomes of waterlily (Nymphaea odorata Ait.) (Lodge, 2005). This wetland is primarily composed of ridges with sawgrass (Cladium jamaicense Crantz) and cattail (Typha domingensis Pers.), sloughs with mostly waterlily, and tree islands. Sawgrass has been the dominant vegetation type in low-nutrient zones of WCA-2A and has been replaced with cattail in those areas associated with phosphorus enrichment and/or fluctuating water levels (Noe et al., 2001). The topography is flat, averaging 3.0 m above sea level (USGS, 2007), and this generates slow sheet flow across the study area, from the northeast to the southwest. Water inflows and outflows in WCA-2A are regulated by the South Florida Water Management District (SFWMD), and are aligned with the water level of adjacent Lake Okeechobee for ecological reasons (Light and Dineen, 1994).

Field Sampling and Soil Carbon Measurement

We collected 108 soil cores in WCA-2A using a 10-cm diameter stainless steel coring tube at two depths (0–10 cm and 10–20 cm) in 2009. We used a stratified random sampling design using historic soil, hydrologic, and ecological sites identified in 2003. A detailed description of the sampling design and the method of soil core collection can be found in Rivero et al. (2007). We could visit only 48 preselected sampling sites, because dense vegetation and a low water table made it difficult to access the sites via airboats, which are a common means of transportation in a wetland. The remaining sampling sites were adjusted to areas as close as possible to the original preselected sampling sites.

Any living organic materials were removed from the samples. All samples were oven dried at 70°C for 72 h to prevent any organic C loss. Weights of the soil samples before and after drying were recorded to measure bulk density (BD) of each sample. Soil samples were ground, ball-milled, and sieved using a 0.45-μm mesh to precede total C (TC) analysis. Soil TC was analyzed with a Shimadzu SSM-5000A TC analyzer (Shimadzu Corporation, Columbia, MD) using 15 to 25 mg of soil and a combustion temperature of 900°C. The amount of TC stored per unit land area (kg m⁻²) for a given depth was calculated using the TC concentrations measured on a mass basis (g kg⁻¹) and the BD measured at each sampling point according to:

\[
TC_{stock} = TC_{conc.} \times \rho \times z
\]

where TC_{stock} is TC stocks (kg m⁻²), TC_{conc.} is TC concentration (g kg⁻¹), \(\rho\) is the BD of the soil sample at each point (g cm⁻³), \(z\) is the depth (10 cm), and \(c\) is a constant for unit conversion (0.01). We performed all statistical analyses, including t-test, using SAS statistical software (SAS Institute, 2010).

Environmental Predictor Variables

We used topographical, hydrological, and lithological properties as environmental predictor variables to develop soil TC stock models (Table 1). Ancillary environmental variables were assembled from various sources. Elevation data surveyed by the USGS using 400-m grid space sampling (±15 cm vertical accuracy) represented topography. Hydrology is one of the most important environmental variables associated with soil nutrient distributions. The surface water inflow and outflow from WCA-2A is regulated by the SFWMD controlling water control structures (WCSSs). There are 11 WCSSs in WCA-2A: S7, S10A, S10C, S10D, S10E, S11A, S11B, S11C, S144, S145, and S146 (Fig. 1). Therefore, a continuous raster-based data layer of distance to WCSSs was derived using the Euclidean distance function to include the effect of hydrology on soil C distributions. The geophysical data, including potassium concentration, \(\gamma\)-ray, bouger gravity anomaly, isostatic residual gravity anomaly, and magnetic anomaly data, represented lithological properties.

In addition, we included various spectral indices derived from satellite images as a proxy of various vegetation specific properties. Spectral indices, such as the normalized difference vegetation index (NDVI), enhance spectral signatures of the Earth’s surface properties, as well as reduce spectral noise. We derived spectral indices from three different RS images: Satellite Pour l’Observation de la Terre (SPOT), Landsat Enhanced Thematic Mapper Plus (ETM+), and Moderate Resolution Imaging...

Journal of Environmental Quality
Spectroradiometer (MODIS) (Table 1). The SPOT images, which have 10-m spatial resolution, were taken in January 2009. The Landsat ETM+ and MODIS images, which have 30- and 250-m spatial resolutions, respectively, were taken in February 2010. The SPOT images were donated by Planet Action, a nonprofit ASTRIUM GEO initiative, and the Landsat ETM+ and MODIS images were released by the USGS. We employed a principal component (PC) analysis to derive PC scores for each of the RS images and utilized a tasseled-cap transformation (TCT) for the Landsat ETM+ image to enhance the spectral signatures of soil brightness, greenness, and wetness (Kauth and Thomas, 1976). Detailed descriptions of the spectral indices and RS images can be found in Kim et al. (2014). All image processes were conducted using ERDAS Imagine 2010 software (ERDAS, 2010). We extracted spectral data from each sensor, spectral indices, PC variables, and TCT variables at each sampling point. All of the geographic information system data processing work was performed using ArcGIS 10 (ArcGIS, 2011).

Data Analysis

We applied random forest (RF) to develop soil C stock models and identify environmental variables that are related to the spatial distribution of soil C stocks. Random forest is a widely used machine-learning method that is commonly used for data exploration and modeling of complex relationships between predictors and target variables. It is a recursive partitioning method that is particularly well-suited to small observation sets and large numbers of predictor variables (Strobl et al., 2009). Random forest is an ensemble of classification or regression trees that are calculated on random subsets of the data (bootstrapped samples) using a randomized subset of predictor variables at each tree (Breiman, 2001). Numerous trees are grown within the algorithm, and one single prediction is given by averaging the trees. Thus, it has shown low bias and low variance (Díaz-Uriarte and de Andrés, 2006). Random forest produces a variable importance to reflect the contribution of each predictor variable to build a forest in RF. The variable importance score for each predictor variable is estimated by looking at how much the prediction error increases when the variable is altered, while all other variables are kept the same (Liaw and Wiener, 2002). Details of the RF algorithm can be found in Breiman (2001).

We developed three different RF models using input variables derived from the SPOT (i.e., RF SPOT model), Landsat ETM+ (i.e., RF ETM+ model), and MODIS images (i.e., RF MODIS model). In this paper, we used an iteration approach to develop a best-fit model for soil C stocks by varying the overall number of trees in the forest ($n_{trees}$) and the number of randomly preselected predictor variables for each split ($m_{try}$) parameter. As a result, we used 1000 of $n_{trees}$ and 7 of $m_{try}$ based on the model’s error rates and the percent of variance explained. Random forest analysis was performed using the “randomForest” package in R (Liaw and Wiener, 2002), and the “raster” package was used to produce prediction maps in the R statistical language (R Development Core Team, 2012). The model performance was assessed using leave-one-out cross-validation using a coefficient of determination ($R^2$) and root mean square prediction error (RMSE).

Results

Descriptive Statistics and Spatial Distribution of Soil Carbon

We present both TC concentration and stock results. Total C stock is influenced by BD and TC concentration measurements; thus, TC stock provides a sensitive measure of the spatial variation and changes in an ecosystem. Descriptive statistics of TC concentrations and stocks are shown in Table 2. The 0- to 10-cm soils showed a mean concentration of 410.5 g kg$^{-1}$, and the 10- to 20-cm soils showed a mean concentration of 432.9 g kg$^{-1}$. The mean of the BD in WCA-2A was 0.11 cm$^{-3}$ for 0- to 10-cm soils and 0.12 cm$^{-3}$ for 10- to 20-cm soils. The observed mean value of soil TC stocks was 4.5 kg m$^{-2}$ for 10 to 20 cm and 5.2 kg m$^{-2}$ for 10 to 20 cm. The 10- to 20-cm layer showed higher ($p < 0.05$) TC concentrations and stocks than the 0- to 10-cm layer. The spatial distributions of soil TC concentrations and stocks measured by depth are shown in Fig. 1. The BD showed a homogeneous distribution throughout the study area; thus, TC concentrations and stocks showed similar spatial patterns, except

<table>
<thead>
<tr>
<th>Property</th>
<th>Variable†</th>
<th>Spatial resolution</th>
<th>Data source‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vegetation</td>
<td>MODIS: reflectance values of blue, red, NIR, and MIR; EVI; MSI; NDVI; TVI; PC 1 to 3</td>
<td>250</td>
<td>USGS/LPDAAC (2010)</td>
</tr>
<tr>
<td></td>
<td>Landsat ETM+: reflectance values of blue, green, red, NIR, MIRs, and PAN; MidIR; MSI; NDVI; NDVI green; NDWI; RSR; SR; TVI; Mean within 3 × 3 cells and 7 × 7 cells for MSI, NDVI, NDVI green, NDWI, RSR, SR, and TVI; PC 1 to 3; TCT 1 to 3</td>
<td>30</td>
<td>USGS/EROSDC (2010)</td>
</tr>
<tr>
<td></td>
<td>SPOT: reflectance values of green, red, NIR, SWIR, and PAN; MSI; NDVI; NDVI green; NDWI; RSR; SR; TVI; Mean within 3 × 3 cells, 9 × 9 cells, and 25 × 25 cells for MSI, NDVI, NDVI green, NDWI, RSR, SR, and TVI; PC 1 to 3</td>
<td>10</td>
<td>SPOT Image Corporation (2009)</td>
</tr>
<tr>
<td>Topography</td>
<td>Elevation</td>
<td>10/30/250</td>
<td>USGS/HAED (2007)</td>
</tr>
<tr>
<td>Hydrology</td>
<td>Distance to water control structures</td>
<td>10/30/250</td>
<td>SFWMD (1997)</td>
</tr>
<tr>
<td>Lithology</td>
<td>Potassium concentration; bouguer gravity; isostatic residual gravity anomaly; magnetic anomaly</td>
<td>2000</td>
<td>USGS/DDS-9 (1999)</td>
</tr>
</tbody>
</table>

† MODIS, Moderate Resolution Imaging Spectroradiometer; NIR, near-infrared; MIR, mid-infrared; EVI, enhanced vegetation index; MSI, moisture stress index; NDVI, normalized difference vegetation index; SR, simple ratio; TVI, transformed vegetation index; PC, principal component; ETM+, Enhanced Thematic Mapper Plus; PAN, panchromatic; MidIR, mid-infrared index; NDVI green, normalized difference vegetation green index; NDWI, normalized difference water index; RSR, reduced simple ratio; TCT, tasseled cap transformation; SPOT, Satellite Oour l’Observation de la Terre; SWIR, shortwave-infrared.

‡ DDS, Digital Data Series; EROSDC, Earth Resources Observation and Science Data Center; HAED, High Accuracy Elevation Data; LPDAAC, Land Processes Distributed Active Archive Center; SFWMD, South Florida Water Management District; USGS, United States Geological Survey.

§ Mean spectral values within neighborhoods (3 × 7, 7 × 9, and 25 × 25 cells, respectively) using the nearest neighbor geospatial method.
in the northeastern and southern portions of the study area. In general, high stocks of TC were found near S10 WCSs, and low stocks were found in the interior marsh.

**Soil Carbon Stock Predictions Using Random Forest**

The statistical assessment of the RF models for soil TC stocks by depth are shown in Table 3. All RF models showed high $R^2$ values ranging from 0.85 to 0.92 and small RMSEs ranging from 0.75 to 0.86 kg m$^{-2}$ for TC stock models. Spatial prediction maps for soil TC stocks are shown in Fig. 2. Soil TC stocks showed higher variability in the 0- to 10-cm than in the 10- to 20-cm depth in prediction maps. All of the RF prediction maps showed low TC stocks in the lower northeastern portion of the study area in 0- to 10-cm soils. For these top soils, the prediction maps of the RF$^\text{SPOT}$ and RF$^\text{ETM+}$ models clearly showed low TC stocks in several tree islands in WCA-2A, which occur on ridges and are less inundated than the surrounding marsh. They also showed TC variations caused by the slough–ridge system, mostly in the

---

**Fig. 1.** Spatial distribution of measured soil total carbon: (a) concentrations (g kg$^{-1}$) and (b) stocks (kg m$^{-2}$) by depth within the Water Conservation Area 2A. Squares are water control structures.

---

**Fig. 2.** Spatial prediction maps of soil TC stocks using the RF$^\text{SPOT}$ and RF$^\text{ETM+}$ models.
important variables in the RFSPOT model to predict soil TC stocks (Table 4). The PC 1 and ‘distance to WCSs’ ranked as the most pared with the 0- to 10-cm soils.

and homogeneous spatial distribution of TC stocks when com-

study area, and the interior of WCA-2A showed relatively low

in 10- to 20-cm depth dominantly occupied the boundary of the

spatial variability than 0- to 10-cm depth soils. High TC stocks

slough–ridge systems. The 10- to 20-cm depth soils showed less

counted as part of the study area in the RFETM+ model (Fig. 2).

the prediction map due to technical sensor problems, were not

expected since the ‘no data’ areas, expressed as blank stripes in

in WCA-2A. The lower TC stocks for the RFETM+ model were

for 0- to 10-cm soils and 2.19 T g (2.19 Mt) for 10- to 20-cm soils

yielded slightly lower amounts of TC stocks: 1.92 T g (1.92 Mt)

R

Table 2. Descriptive statistics for soil total carbon concentrations and stocks observed in Water Conservation Area 2A (n = 108).

<table>
<thead>
<tr>
<th>Property</th>
<th>Depth</th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Median</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentrations, g kg⁻¹</td>
<td>0–10</td>
<td>410.5</td>
<td>262.8</td>
<td>492.3</td>
<td>413.2</td>
<td>35.6</td>
</tr>
<tr>
<td></td>
<td>10–20</td>
<td>432.9</td>
<td>325.5</td>
<td>486.7</td>
<td>438.1</td>
<td>32.5</td>
</tr>
<tr>
<td>Stocks, kg m⁻²</td>
<td>0–10</td>
<td>4.5</td>
<td>2.2</td>
<td>10.8</td>
<td>4.0</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>10–20</td>
<td>5.2</td>
<td>1.3</td>
<td>11.2</td>
<td>4.7</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Table 3. Summary of model performance assessment for soil total carbon stocks (kg m⁻²).

<table>
<thead>
<tr>
<th>Depth</th>
<th>Statistical measure†</th>
<th>RFSPOT‡</th>
<th>RFETM‡</th>
<th>RFMODIS‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–10</td>
<td>$R^2$</td>
<td>0.85</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.75</td>
<td>0.77</td>
<td>0.80</td>
</tr>
<tr>
<td>10–20</td>
<td>$R^2$</td>
<td>0.92</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.79</td>
<td>0.78</td>
<td>0.86</td>
</tr>
</tbody>
</table>

† $R^2$, coefficient of determination; RMSE, root mean squared error.
‡ RFSPOT, RFETM+, and RFMODIS, random forest models using Satellite Pour l’Observation de la Terre (SPOT), Landsat Enhanced Thematic Mapper Plus (ETM+), and Moderate Resolution Imaging Spectroradiometer (MODIS) image-derived input variables, respectively.

central and southern portion of WCA-2A. However, the prediction map of the RF MODIS model showed neither tree islands nor slough–ridge systems. The 10- to 20-cm depth soils showed less spatial variability than 0- to 10-cm depth soils. High TC stocks in 10- to 20-cm depth dominantly occupied the boundary of the study area, and the interior of WCA-2A showed relatively low and homogeneous spatial distribution of TC stocks when compared with the 0- to 10-cm soils.

Impacts of each predictor variable varied among the RF models (Table 4). The PC 1 and ‘distance to WCSs’ ranked as the most important variables in the RF SPOT model to predict soil TC stocks for both 0 to 10 and 10 to 20 cm. The PC 1 in the RF SPOT model was derived mostly from near-infrared (NIR) and shortwave-infrared (SWIR) bands of the SPOT image in the PC analysis. This variable explained about 75% of spectral signatures of WCA-2A, and it represented different land covers (i.e., vegetation and water) of the study area. For the RF ETM+ model, the mid-infrared index (MidIR) and TCT variable 2 (TCT 2) ranked as the best variables to predict soil TC stocks for 0 to 10 and 10 to 20 cm, respectively. For the RF MODIS model, ‘distance to WCSs’ and the NIR reflectance ranked high to infer on soil TC stocks for 0 to 10 and 10 to 20 cm, respectively. All three TC prediction models were dominated by strong, spectral-based predictors (Table 4).

### Total Carbon Stocks in Water Conservation Area 2A

Here, we focus on spatially-explicit soil C stock assessment, whereas Kim et al. (2014) discussed the effects of spatial resolutions of RS images in more detail. The RF SPOT and RF MODIS models showed similar TC stocks, which are 1.96 Tg (1.96 Mt) for 0- to 10-cm soils and 2.24 and 2.23 Tg (2.24 and 2.23 Mt) for 10- to 20-cm soils, respectively (Table 5). The RF ETM+ model yielded slightly lower amounts of TC stocks: 1.92 Tg (1.92 Mt) for 0- to 10-cm soils and 2.19 Tg (2.19 Mt) for 10- to 20-cm soils in WCA-2A. The lower TC stocks for the RF ETM+ model were expected since the ‘no data’ areas, expressed as blank stripes in the prediction map due to technical sensor problems, were not counted as part of the study area in the RF ETM+ model (Fig. 2).

Overall, all of the RF models yielded similar TC stocks. For comparison, TC stocks using observed mean and median values weighted by the total area of WCA-2A were estimated, and they showed slightly lower TC stocks than the RF models (Table 5). Based on the prediction results, about 4.2 Tg (±0.05, 4.2 Mt) of C was stored in the top 20 cm of the study area, and it yielded about 9.5 kg TC stocks per unit area (m⁻²) in the top 20 cm in WCA-2A.

### Discussion

### Spatial Distribution of Soil Carbon

Total C concentrations in both soil layers were lower than in previously documented studies conducted in WCA-2A (DeBusk et al., 1994; Rivero et al., 2007; Wright and Reddy, 2007). All of the previous studies used the same methods and procedures to measure TC concentrations as were used in this study. Thus, differences in methods do not explain different TC concentrations among the studies conducted in WCA-2A. However, these studies differed in terms of sampling site selection, sampling density, and seasonality of sampling. The studies also showed increasing TC concentrations with depth, and the same trend was observed in WCA-1 (Corstanje et al., 2006), the ENP area (Osborne et al., 2011), and WCA-3B (Bruland et al., 2006), all located within the Greater Everglades. For TC stocks in WCA-2A, the 10- to 20-cm depth showed significantly higher (p < 0.05) stocks than the 0- to 10-cm depth. No previous studies have reported the vertical distribution of TC stocks in the Everglades region. In contrast to our results, Wannen et al. (2012) reported decreasing soil organic C stocks with depth in predominantly herb-covered wetlands in Brazil. Most studies on soil C stocks conducted in other land uses have observed decreasing stock values with depth. For instance, Grimm et al. (2008) found the highest stocks in the upper 10-cm soils in a tropical forest located in Panama. Vasques et al. (2010) observed decreasing soil TC along soil profiles (down to 180-cm depth) in the Santa Fe River watershed in Florida, which also contained several wetlands.

### Soil Carbon Stock Predictions Using Random Forest

The high $R^2$ and low RMSE values of the RF models showed great confidence to assess soil TC stocks in WCA-2A using various environmental predictor variables. Random forest has shown high performance in predicting soil C in other ecosystems (Grimm et al., 2008). For instance, Wiesmeier et al. (2011) predicted soil organic C stocks up to a 1-m depth in a semiarid ecosystem in China, and the RF model yielded an $R^2$ of 0.74.

The RF prediction maps for 0- to 10-cm soils showed consistently low TC stocks in the lower northeastern portion of the study area, which spatially coincide with soils enriched in phosphorus (Kim et al., 2014). One possible explanation is that high C losses have occurred within this local region as a result of the phosphorus enriched inflow from the WCSs. Typically,
This area has shown high soil total phosphorus concentrations and dense vegetation communities dominated by cattail because of phosphorus-enriched inflow from the adjacent Everglades Agricultural Area (Grunwald et al., 2004). It is known that cattail detritus are better electron donors in methane production than sawgrass because of their cellulose and lignin contents. This was underpinned by Gerard and Chanton (1993) and Wright and Reddy (2007) showing significantly increased methane production in phosphorus-enriched areas of WCA-2A. Pinney et al. (2000) also reported high methane release within cattail communities in a constructed wetland in Arizona. This loss of C might produce profoundly low C stocks in phosphorus-enriched areas of WCA-2A.
areas. Our prediction maps of the $RF_{\text{SPOT}}$ and $RF_{\text{ETM+}}$ models clearly showed the effect of the landforms (i.e., tree islands and slough–ridge systems) on soil C stocks, especially in the central and southern area. Although we did not collect any samples from tree islands, our prediction results showed some agreement with previously observed TC values in tree islands (Hanan and Ross, 2010).

Less spatial variability of soil TC stocks in the 10- to 20-cm layer was expected, since the organic matter in subsurface and deep soils are highly decomposed and, as a result, they have a more homogeneous distribution than surface soils. The relatively high TC stocks in boundary areas compared with the interior marsh in 10- to 20-cm soils could be explained by different hydroperiods. The boundary areas usually have deep water depth because of canals, and this creates longer hydroperiods than found in the interior of the marsh. The soils under long hydroperiods have more of a chance to retain organic matter and thus accrete C in soils (Osborne et al., 2011; Wang et al., 2011). Our results suggest that the empirical RF models with RS image-derived spectral indices are robust, not only to predict soil TC stocks in the top soil, but also to predict subsoil C stocks, although RS images derived from passive sensors can only observe the earth's surface features.

We could not find one major common environmental predictor variable that explained the spatial distribution of C stocks in WCA-2A, as indicated by variable importance rankings (Table 4). Most of the highly ranked importance variables, however, carried information relating to hydrology (e.g., PC 1, ‘distance to WCSs,’ and MidIR) and/or vegetation (e.g., TCT 2 and NIR). For instance, ‘distance to WCSs’ that ranked high in the $RF_{\text{SPOT}}$ and $RF_{\text{MODIS}}$ models contained combined effects of nutrient loadings and inflow across the study area, and the MidIR showed strong correlation with soil moisture (Musick and Pelletier, 1988). The NIR is very responsive to the amount of vegetation biomass present, and TCT 2 contains greenness (vegetation) information (Kauth and Thomas, 1976). Other studies reported topographical factors as important variables for soil C predictions (Minasny et al., 2006; Grimm et al., 2008; Wiesmeier et al., 2011). However, our results showed a low ranking of elevation in the prediction models. These findings suggest that elevation is less influential compared with other environmental variables explaining the distribution of C stocks in our study area, which is dominated by smooth and/or flat topography and macrophytes commonly found in subtropical wetlands in Florida.

Total Carbon Stocks in Water Conservation Area 2A

Although we only estimated TC stocks in the top 20-cm soils, the contribution of deep soils to global soil C stocks has been proposed to be much greater than surface soils. For instance, Jobbágy and Jackson (2000) estimated that more than 50 to 70% of soil organic C is stored at depths greater than 20 cm. Schmidt et al. (2011) estimated that more than half of the global soil C stocks are stored in deep soils. This might be especially true in wetlands consisting of deep layers of peat. Since the soils in our study area were classified as Haplosaprists, the O horizon contained most of the soil organic matter and, typically, organic matter contains about 45 to 50% C (Kaynarli et al., 2010). Consequently, the O horizon contained the majority of soil organic C in WCA-2A. Bhatti and Tarnocai (2009) reported that only 14% of soil C stocks were stored in the top 30 cm of organic soils in Canada. Therefore, the estimated 4.2 Tg (4.2 Mt) of C (0- to 20-cm depth) in our study might be only a small portion of the total amount of C stored in this wetland, and a substantially greater amount of C is expected to be stored below 20 cm. The calculated median TC stocks per unit area (i.e., 9.5 kg m$^{-2}$) indicates the C-richness of this wetland. It was higher than 8.9 kg m$^{-2}$, which was the estimated mean of TC stocks in 0- to 100-cm depth of wetland soils in Pennsylvania (Kumar et al., 2012). It was also much higher than 7.3 kg m$^{-2}$, which is the estimated median of TC stocks in all of the US wetlands in the top 20 cm using the State Soil Geographical database (Guo et al., 2006). Our TC stock estimate was also higher than the average C stocks per unit area of wetlands located in other countries (Bernal and Mitsch, 2008; Wiesmeier et al., 2011; Wantzen et al., 2012). Furthermore, our topsoil TC stock estimate was much higher compared with C stocks in upland soils. For instance, Vasques et al. (2010) estimated an average C stock of 7.6 kg m$^{-2}$ in the top 30-cm soils in northern Florida. The WCA-2A only accounts for 6.5% of the total area of the Everglades region, and previous studies reported higher soil organic matter and soil organic C concentrations in other units, such as WCA-1 and WCA-3 (Corstanje et al., 2006; Osborne et al., 2011). The Greater Everglades might therefore store enormous amounts of C in soils, and it suggests that this C-rich wetland needs more attention when estimating the global C budget.

Reddy et al. (1993) reported a soil C sequestration rate of 86 to 387 g C m$^{-2}$ yr$^{-1}$ in WCA-2A that was much greater than the estimated average soil C sequestration rate of North American wetlands (29 g C m$^{-2}$ yr$^{-1}$; Gorham, 1991). The high values of C sequestration are related to hydrology (Marín-Muñiz et al., 2014), and water level fluctuations can significantly alter the C balance in wetlands (Couwenberg et al., 2010). Water levels (i.e., inflow rate at WCSs) of WCA-2A are controlled by the SFWMD to prevent floods and droughts and to provide ecosystem services. South Florida has experienced a prolonged drought since 2006. Low water levels have exposed highly organic soils of the wetland to the atmosphere during the drought, and it likely released a considerable amount of C stored in this wetland's soils. Extreme climate events (i.e., prolonged drought) pose the risk that this wetland loses its role as a C sink. South Florida has experienced increasing temperatures and decreasing precipitation over the past 10 yr. Moreover, climate change projections for southern Florida suggest an overall drier climate, with possible intensification of prolonged drought conditions (Ingram et al., 2013) that would profoundly impact soils in WCA-2A and possibly other wetlands in southern Florida. Hydrologic modeling with future climate scenarios showed significant dry outs of the Everglades with decreased rainfall and increased evapotranspiration (Obeysekera et al., 2015), possibly forcing the Everglades to shift from a sink to a source of C.

Wetland soils sequester a much greater amount of C than other ecosystems (Guo et al., 2006). At the same time, C loss from wetland soils is much faster than from other soil types, and it increases with the previous organic C content (Bellamy et al., 2005). Ross et al. (2013) found the largest net C loss in Histosols (Saprists) in the St. Johns River Basin in Florida. Since soils in WCA-2A are dominated by high organic matter (Saprists), the area is at risk to lose large amounts of C under prolonged drought conditions, potentially intensified through global climate warming. Numerous studies have shown that a
wetland acts as a C sink rather than as a C source when it is managed appropriately (Armentano and Menges, 1986; Lal, 2003; Koelbener et al., 2010). Understanding the vulnerability of the water management system in WCA-2A is critical to preserve this C-rich ecosystem.

Conclusions
Our study showed successful model development to predict soil TC stocks in the 0- to 10-cm and 10- to 20-cm depths of WCA-2A using RS image-derived spectral indices and environmental variables. Our results showed that WCA-2A stored a tremendous amount of C in soils. Although we could not find a specific environmental or spectral variable that explains the spatial distribution of soil TC stocks in the area throughout depths, several variables related to hydrology ranked high in the variable importance of the prediction models. This suggests that the hydrology controlled by the SFWMD is critical to affect C cycling in the region and globally. Our results also suggest that understanding the spatial distribution of soil TC stocks and appropriate management of the wetland is critical to maintain its ability to store C. Large changes in temperature and precipitation in the Everglades region are predicted to occur over the next several decades. Thus, sequentially reproducible assessment methods are necessary for measuring and monitoring the impacts of those changes in C-storage capacity of the wetland. Remote sensing-supported DSMM suits the need well, as it provides a methodology to access the amount of C stored in the wetlands.

Acknowledgments
Funding for this project was provided by the Cooperative Ecosystem Studies Unit—Natural Resources Conservation Service and the GIS–Pedometrics research group, Soil and Water Science Dep., Univ. of Florida. The SPOT image was provided by Planet Action. We thank Dr. Todd Z. Osborne, Greg Brannon, Howard Yamatani, Marig Figueroa, S. Matt Norton, and Kevin Ratkis for field sampling and Tom Weber for coordination.

References

ERDAS. 2010. ERDAS Imagine software. Earth Resource Data Analysis System Inc., Atlanta, GA

Journal of Environmental Quality
1917
