Modeling soil parameters using hyperspectral image reflectance in subtropical coastal wetlands

Naveen J.P. Anne\textsuperscript{a}, Amr H. Abd-Elrahman\textsuperscript{a,*}, David B. Lewis\textsuperscript{b}, Nicole A. Hewitt\textsuperscript{a}

\textsuperscript{a} School of Forest Resources and Conservation – Geomatics, University of Florida, Florida, United States
\textsuperscript{b} Department of Integrative Biology, University of South Florida, United States

\textbf{A R T I C L E   I N F O}

\textbf{Article history:}
Received 25 December 2013
Accepted 9 April 2014

\textbf{Keywords:}
Hyperspectral remote sensing
Coastal wetlands
Soil properties
Particulate organic matter
Labile carbon
Labile nitrogen

\textbf{A B S T R A C T}

Developing spectral models of soil properties is an important frontier in remote sensing and soil science. Several studies have focused on modeling soil properties such as total pools of soil organic matter and carbon in bare soils. We extended this effort to model soil parameters in areas densely covered with coastal vegetation. Moreover, we investigated soil properties indicative of soil functions such as nutrient and organic matter turnover and storage. These properties include the partitioning of mineral and organic soil between particulate (>53 μm) and fine size classes, and the partitioning of soil carbon and nitrogen pools between stable and labile fractions. Soil samples were obtained from Avicennia germinans mangrove forest and Juncus roemerianus salt marsh plots on the west coast of central Florida. Spectra corresponding to field plot locations from Hyperion hyperspectral image were extracted and analyzed. The spectral information was regressed against the soil variables to determine the best single bands and optimal band combinations for the simple ratio (SR) and normalized difference index (NDI) indices. The regression analysis yielded levels of correlation for soil variables with $R^2$ values ranging from 0.21 to 0.47 for best individual bands, 0.28 to 0.81 for two-band indices, and 0.53 to 0.96 for partial least-squares (PLS) regressions for the Hyperion image data. Spectral models using Hyperion data adequately (RPD > 1.4) predicted particulate organic matter (POM), silt + clay, labile carbon (C), and labile nitrogen (N) (where RPD = ratio of standard deviation to root mean square error of cross-validation [RMSECV]). The SR (0.53 μm, 2.11 μm) model of labile N with $R^2 = 0.81$, RMSECV = 0.28, and RPD = 1.94 produced the best results in this study. Our results provide optimism that remote-sensing spectral models can successfully predict soil properties indicative of ecosystem nutrient and organic matter turnover and storage, and do so in areas with dense canopy cover.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Coasts are subject to devastating storms and frequent tidal exchange (Cahoon, 2006), making coastal wetlands vital to preventing shoreline erosion by providing sediment stabilization and water storage (Hardisky et al., 1986). In addition, coastal wetlands process organic and chemical wastes and reduce sediments in water (Rao et al., 1999). In turn, soil quality and composition have a direct effect upon the health of vegetation, especially in wetland environments (Ehrenfeld et al., 2005), and maintaining this reciprocal relationship is essential to preserving coastal habitats. For instance, coastal soils hold large concentrations of organic matter that supplies the N needed to support mangrove forest and salt marsh primary production (Nedwell et al., 1994; Anderson et al., 1997; Alongi et al., 2002), which in turn promotes soil accretion through organic matter production and sediment trapping (Morris et al., 2002). This feedback is critical for maintaining wetland elevation relative to rising sea levels (Kirwan and Mudd, 2012; Morris et al., 2012). Soils are complex and dynamic, both temporally and spatially, requiring numerous physical, chemical, and biological determinants for soil quality assessment. Soil sampling, from specimen collection to the generation of quantitative data, consumes a tremendous amount of time and requires delicate lab procedures.

The premise of estimating soil components using spectral analysis under laboratory conditions was heavily tested along the past few decades. For example, in agricultural settings, soil organic matter (SOM) models were defined through spectral reflectance measured in the laboratory for black soil (Liu et al., 2009), and litchi orchid (Li et al., 2012). Similarly, sand, clay (Minasny et al., 2008; Summers et al., 2011), total C, total N (Morra et al., 1991), and soil
moisture (Dalal and Henry, 1986) were tested in laboratory settings. Shepherd and Walsh (2002) tested topsoil (0- to 15-cm depth) organic C concentration, clay content and sand content through diffuse reflectance spectroscopy (0.35–2.5 μm) on a diverse sample that was acquired from a wide variety of landscape positions, parent materials and landscapes. Vitagesu et al. (2011) tested modeling pure clay material using laboratory based spectral data (2.5–14 μm) and partial least squares (PLS) regression and attained much of the variation in their components. In these studies, strong correlations between spectral reflectance and the field samples were reported. For example, absorption regions of C and N bonds of near-infrared reflectance spectroscopy (NIRS) correlated with soil C and N fractions (Morra et al., 1991). Near-infrared bands attributed to the bending and stretching of O–H bonds in lattice mineral and water molecules were associated with soil clay content (Summers et al., 2011).

Brown et al. (2006) found strong relationships between VNIR and soil organic C, and clay content. They identified the 0.54, 0.55, and 1.91 μm wavelengths, sensitive to H2O, as important wavelengths in soil organic C estimation. Information in the 2.0–2.5 μm was found important due to e.g. C–O and C–H bond absorptions. Dalal and Henry (1986) concluded that absorption along the near-infrared (NIR) spectra was correlated with moisture content in soil. Summers et al. (2011), and Chang et al. (2001) concluded that the organic matter content of the soil was inversely proportional to albedo along the visible near-infrared (VNIR) region of the spectrum. Bands that significantly correlated with SOM samples were associated to regions in the spectra sensitive to the C–H, N–H, and O–H bonds (Li et al., 2012). These studies used in situ or laboratory spectroscopy of bare soils regressed against the laboratory based physically and chemically derived soil properties.

Technological advances in remote sensing have allowed for the development of cost-efficient and effective alternative methods of soil assessment (Mulder et al., 2011). The transition from lab to remote sensing analysis of soils involves accounting for the effects of atmospheric influences, geometric distortions, spatial resolution, and scale (Mulder et al., 2011). Several surface soil properties including organic matter were modeled from airborne hyperspectral imagery (Hbirkou et al., 2012; Ben-Dor et al., 2002) and aerial color photographs (López-Granados et al., 2005). Airborne hyperspectral imagery was used to build SOM models for agricultural clay-loam soil (Uno et al., 2005). Such models were successfully constructed using field samples and aerial hyperspectral data (HyMap and Airborne Visible/Infra-Red Imaging Spectrometer) of bare soil fields for estimating SOM, sand, silt, clay, and other soil properties (Palacios-Orueta and Ustin, 1998; Selige et al., 2006).

Chabrillat et al. (2002) showed the use of AVIRIS and Hyperspectral Mapper (HyMap) imagery and matched filtering algorithm for successful mapping of exposed clay minerals. Combined sampling of dry bare ground and pasture was used in modeling soil organic carbon using the spaceborne measurements and Hyperion hyperspectral imagery (Lu et al., 2012; Gomez et al., 2008). Mulder et al. (2013) aimed at the characterizing and improving the mapping of mineral variability at a regional scale from remote sensing imagery on a diverse lithological setting that included sedimentary, igneous, and metamorphic rock types. ASTER, multiple linear regressions (MLR), and different smoothing techniques were used to evaluate clay mineral and attained moderate R² values of 0.57 and 0.45. Shi et al. (2014) used HyMap airborne hyperspectral imagery and field samples to map soil acidity in coastal areas. Their study used PLS regression and mineral mapping as an indicative of soil acidity.

Although numerous studies have focused on modeling soil properties from remote sensing technologies either directly from bare soil, or by inferring soil properties through vegetation cover (Huete, 2005; Kooistra et al., 2004). However, utilization of remote sensing technology in quantification of under-canopy soil properties remains limited. Ben-Dor et al. (2002) developed methods to quantify under-canopy soils properties, but did so with interpolation techniques such as kriging (López-Granados et al., 2005) that were based on models built with open-surface soil samples. Spatial interpolations, however, present accuracy problems in mapping applications (Selige et al., 2006). In other studies, soil properties were mapped from partially vegetated fields through spectral unmixing of hyperspectral data for estimating clay (Querghemmi et al., 2011) and soil organic carbon (Bartholomeus et al., 2011). The research on inferring soil properties through vegetation cover is in its infancy, as only a few studies have investigated the relationship between soil properties and vegetation reflectance (Kooistra et al., 2004; Piekarczyk et al., 2012). Kooistra et al. (2004) estimated substrate (vegetation covered) SOM and soil moisture in addition to a few other properties through field measured hyperspectral vegetation reflectance data. Gomez et al. (2008) focused on establishing the relationship between Hyperion spectra and in situ soil samples of bare soil and pasture. Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) were used to infer few substrate soil properties, including reflectance through vegetation (Piekarczyk et al., 2012).

While the application of remote sensing to soils continues to grow in terms of context (e.g., modeling sub-canopy soils), it is also at its infancy in terms of information value for ecosystem function. Remotely sensed soil characteristics evaluated to date (cited above) are important components of ecosystem structure, and are relevant to global change issues (e.g., total soil C storage). They are not, however, the most important soil characteristics for functions such as nutrient recycling, primary production, and soil CO₂ emissions. For these functions, soil indicators of nutrient recycling and organic matter recalcitrance would be more useful. Such indicators include (a) the fractionation of mineral and organic soil between particulate (>53 μm) and fine size classes, as particle size can regulate the availability of soil organic matter to microbial decomposers and can mediate rates of N immobilization in soil (Sollins et al., 1996; Stewart et al., 2007; Castellano et al., 2013), and (b) the partitioning of soil organic carbon and nitrogen between stable and labile (readily-mineralized) fractions, which is thought to regulate the capacity of soil to retain these elements (Kaye et al., 2002). Labile stores of soil carbon and nitrogen are readily transformed to CO₂ and nitrogen oxides, which can pose air and water quality problems. We are aware of no studies that evaluate whether these functional soil characteristics correlate with spectral reflectance.

The goal of this study is to develop statistical models that estimate soil parameters related to nutrient and organic matter storage and turnover from the Hyperion and Thematic Mapper satellite imagery, providing a practical method for remotely monitoring soil composition in coastal wetland environments. We hypothesize that soil characteristics are affected by plant cover, which in turn correlates with image spectra. In this context, this study encompassed areas of dense vegetation to investigate the levels of correlation between hyperspectral/multispectral band reflectance and eighteen soil parameters such as particulate organic matter (POM), mineral associated organic matter (MAOM), labile carbon (labile C), and labile nitrogen (labile N), which require significant resources and extended field sampling procedures. To our knowledge, many of the soil properties investigated in this study have not been previously studied through the vegetation reflectance technique.

2. Materials and methods

2.1. Soil sampling

Soils were sampled from three intertidal sites (Fig. 1) at intervals of about 4.4 km along the coast of west-central peninsular Florida,
From each composite, subsamples were apportioned to determine soil moisture, mineral soil and organic matter particle size distributions, and stable and labile fractions of soil organic carbon and nitrogen. Soil moisture was determined as the fraction of soil mass lost on drying at 60 °C to constant mass, and soil bulk density was calculated as dry soil mass per unit core volume. Total mineral soil content was determined as the fraction of total soil dry mass remaining after ignition of the soil at 550 °C for 4 h, while total SOM was determined as the fraction of soil mass lost on ignition. Using a separate subsample, sand content was calculated as the mass of ash-free (post-ignition) mineral soil retained on a 53-μm sieve. POM was calculated as the mass of >53 μm soil particles lost on ignition. Prior to wet sieving the subsamples used in sand and POM determination, soils were dispersed by shaking for 2 h in a solution of sodium hexametaphosphate (HMP, or (Na(PO3)2)h). The HMP was prepared at a concentration of 3% by mass in deionized water, and the soil was shaken in a slurry of 3 ml HMP per g dry soil (Kettler et al., 2001). MAOM was computed as SOM – POM, while the silt + clay content of the mineral soil was computed as total mineral soil content – sand.

Incubation assays were conducted to partition total soil organic carbon and nitrogen into labile and stable fractions. Sixty grams of soil from each plot-level composite sample were placed in a filtration apparatus, and incubated in the dark at room temperature (~25 °C) and field capacity moisture for 365 days. Labile C and N were quantified as the masses of carbon and nitrogen mineralized from the incubated soil during the 365-d period. To quantify labile C, we estimated CO2 efflux per day (µmol CO2 g⁻¹ soil d⁻¹) on 29 occasions during the incubation. These per-day flux rates were then integrated via interpolation over the year. Observations of per-day flux were obtained by sealing samples in airtight chambers, and then quantifying the rate of headspace CO2 accumulation by injecting headspace air samples into an infrared gas analyzer (PP Systems EGM-4 IRGA; Amesburg, MA, USA). To estimate labile N, each sample was leached with a C- and N-free solution (Nadelhoffer, 1990) on 19 occasions during the 365-day incubation, and then extracted in 0.5 M potassium sulfate (K2SO4) at the end of the incubation in order to remove any residual mineral N. Labile N was then quantified as the total mass of mineral N leached and extracted. Mineral N in leached and extracted solutions was detected colorimetrically as nitrate-N + nitrite-N (NO3−−N + NO2−−N) and ammonium-N (NH4−−N) on a microplate spectrophotometer (Biotek Epoch; Winooski, VT, USA) (Hood-Nowotny et al., 2010). This incubation technique includes new C and N inputs and persistently leaches mineral N, forcing microorganisms to meet demand by mineralizing existing pools or carbon and nitrogen. It thereby directly assesses the biological availability (liability) of soil organic C and N pools present at the time of soil sampling (Nadelhoffer, 1990; Robertson and Paul, 2000; Kaye et al., 2002). Total soil C and N were determined from dried, milled subsamples subjected to elemental analysis (ECS 4010, Costech, Inc., Valencia, CA, USA). Stable C and N were determined as total C and N minus labile pools.

### 2.2 Remote sensing

Two remote sensing images were used in this study. The first image is a Hyperion hyperspectral image, captured on September 29th, 2010 by the Earth Observing-1 (EO-1) satellite (Fig. 1) from an altitude of 705 km. Scene center location of the Hyperion image is 28.0598° N latitude and 82.8218° W longitude. Geographic coordinates (World Geodetic Datum [WGS] 1984) of the bounding rectangle enclosing the in situ plots are 28.3369° N, 82.7657° W (upper left) and 28.2445° N, 82.7054° W (lower right). The image was captured around 16:03:05 GMT and contains 10–19% cloud cover. This image includes 242 spectral bands (0.4–2.5 μm) with a 30 m spatial resolution. The Level 1Gst (radiometrically corrected
and resampled for geometric rectification and registration) product was downloaded from the United States Geological Survey website (USGS, 2010) in Geographic Tagged Image File Format (GeoTIFF) format. The LiGst data was provided in 16-bit radiance values and has the parallax error from topographic relief corrected with the use of a digital elevation model (DEM) generated from the Shuttle Radar Topography Mission (SRTM) and other necessary elevation data (USGS, 2010).

2.3. Image atmospheric correction, noise reduction and geometric correction

Remote sensing images are affected by environmental factors such as aerosol, water vapor, and atmospheric gases. Elimination of atmospheric effects, which includes the conversion of radiance to reflectance values, is a prerequisite for quantitative remote sensing analysis (Gao et al., 2009). Weather records from the National Oceanic and Atmospheric Administration (NOAA, 2012) report the local temperature (83°F with 72% humidity) at the time the Hyperion data was captured. The Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercube (FLAASH) algorithm, based on the MODTRAN4 radiative transfer model, removes atmospheric effects mentioned above and outputs at-surface-spectral-reflectance values (Felde et al., 2003). The FLAASH algorithm was applied to the images with a tropical atmospheric model and maritime aerosol model using Environment for Visualizing Images (ENVI 4.8) software. For the Hyperion image, post atmospheric correction, 106 of 242 bands were eliminated owing to either low signal-to-noise ratio (<0.41 μm and >2.4 μm), water absorption (at 1.40 and 1.9 μm) or redundant bands resulting from the overlap of the visible near-infrared (VNIR) and shortwave-infrared (SWIR) bands of the Hyperion sensor.

Hyperspectral imagery acquired from the Hyperion sensor is prone to noise due to its pushbroom scanning system. Noise in the image was minimized using the minimum noise fraction (MNF) algorithm. The MNF filtering algorithm separates noise and the signal in the imagery (Pande-Chhetri and Abd-Elrahman, 2013) and outputs coherent data with reduced noise (Jensen, 2005). The ENVI 4.8 software was used to perform the MNF transformation on the atmospherically rectified reflectance data. The first several bands of the MNF transformed image with higher signal-to-noise ratio (eigenvalue of 99.7% or higher) were inverse transformed to the image domain. This resulted in a noise-reduced reflectance image.

The processed Hyperion reflectance image was based on the WGS-84 datum with a Universal Transverse Mercator (UTM) projection (Zone 17 North). The image covering the study area has a large amount of open water and vast stretches of undeveloped coastline, which does not provide a sufficient number of strong control features equally distributed across the entire image. This proved to be problematic when attempting to correct the image geometrically. In order to circumvent this issue, the image was subset into three small images centered on each survey site, and then each subset image was geometrically corrected separately. All subsets were registered to aerial digital orthophoto quarter quadrangle (DOQQ) images with 1 m pixel resolution provided by the Land Boundary Information System (LABINS, 2012). A total of 34 ground control points (GCPs) were used to rectify these subsets. A first degree polynomial warping method was used to produce subsets with minimal geometric errors. The root mean square error (RMSE) values of the image georeferencing process were 0.48, 0.41 and 0.48 pixels for the 3 image subsets.

2.4. Band and spectral indices regression

After atmospheric corrections, pixel spectral values of the Hyperion imagery at the field plot locations were extracted. The Hyperion reflectance spectra curves corresponding to the 18 in situ observations (plots) are shown in Fig. 2. Values extracted from each of the processed Hyperion image bands (136 of 242) were linearly regressed against the field measured soil properties. The coefficient of determination (R²), root mean square error of cross-validation (RMSECV), and the ratio of the standard deviation to the RMSECV (RPD) of the models were calculated for these linear regression models. The RMSECV was calculated based on the leave-one-out algorithm (Martens and Dardenne, 1998), which systematically ignores one observation from the total number of observations, derives the model, predicts the value of the ignored observation from the derived model, and computes the residual of the singled-out observation. The process is repeated for all the observations in the sample. The summation of the squares of the residuals is termed the predicted sum of squared residuals (PRESS) statistics. The square root of the PRESS value divided by the total number of observations in the sample is the RMSECV.

The RPD metric has been used extensively for assessing the accuracy of soil properties prediction models (Chang et al., 2001; Dunn et al., 2002; Gomez et al., 2008; Summers et al., 2011). Its interpretation differs among disciplines. Many studies interpreted RPD values of 2 or more as an indication of model ability to predict soil properties; RPD values between 1.4 and 2.0 indicate acceptable models that have room for improvement; RPD values below 1.4 are an indication of decreasing reliability of the model; and finally RPD values below 1 indicate that the mean of the observations would be better than the predictors (Chang et al., 2001; Summers et al., 2011).

In addition to individual band regression, the simple ratio (SR), \( \lambda_1/\lambda_2 \) (Baret and Goujon, 1991), and normalized difference index (NDI) \( (\lambda_1 - \lambda_2) / (\lambda_1 + \lambda_2) \) (Jackson and Huete, 1991), were two-band spectral indices used in the regression analysis of soil properties. The SR and NDI were tested on all band combinations of the Hyperion image. These spectral indices are widely used in estimating biophysical parameters of vegetation, and our objective was to understand the agreement between in situ soil properties and remote sensing data of vegetation cover.

2.5. Partial least-squares regression

Partial least-squares (PLS) regression is a multivariate analysis technique used when predictors (factors or independent variables) are large in number and are highly collinear (Wold et al., 2001). Partial least-squares regression is widely used in soil property studies at different scales such as laboratory and in situ spectral analysis in addition to aerial and space-borne hyperspectral image analysis (Dunn et al., 2002; Gomez et al., 2008; Summers et al., 2011; Oertherhemmi et al., 2011). The PLS method reduces the entire reflectance spectra to a small number of relevant factors and regresses them to the dependent variable (Gomez et al., 2008). Partial least squares regression decomposes both the predictor (Hyperion spectral bands) and response variables (soil properties) and identifies latent vectors that maximize the co-variability between them (Wold et al., 2001; Li, 2008; Wang et al., 2011): it does not, however, examine or explain the relationship between variables (Tobias, 1995).

Eqs. (1) and (2) show the decomposition of the X (predictor) and Y (response) variables.

\[
X = TP^T + E
\]  
\[
Y = UQ^T + F
\]

where \( T \) and \( U \) are \( n \times m \) (\( n = \) number of observations; \( m = \) number of predictors) score matrices, \( P (p \times m) \) (\( p = \) number of predictors; \( m = \) number of components) and \( Q (r \times m) \) (\( r = \) number of responses and \( m = \) number of components) are loadings matrices, and \( E \) and \( F \) are the residual matrices of \( X \) and \( Y \) respectively. Prime (\( P' \) or \( Q' \))
Fig. 2. Hyperion reflectance spectra of all 18 plots. The spectral curves of the mangrove plots are represented as dotted (red) curves, and Juncus as dashed (blue) curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

indicates the matrix transposition. The X-scores, T, is a matrix of uncorrelated columns (latent vectors) that are linear combinations of Hyperion spectra (predictor variables). The Y-scores matrix (U) portrays the corresponding variation in the response variable (e.g. POM, MOAM, etc.). The X-loadings matrix (P), where the loading values represent the significance of the corresponding predictor of the mth component, links the Hyperion spectra (predictor variable) to the X-scores. Y-loadings (Q) relate the response variable with the Y-scores. Subsequently, PLS aims to minimize the norm of F while maximizing the co-variability between the response and predictors by the multiple linear regression between the score matrices U and T.

\[ U = BT \] (3)

B is an \( n \times n \) regression coefficient matrix that emerges from the regression.

In this study, we used the reflectance spectra (after applying the atmospheric correction and minimum noise fraction de-noising as explained in Section 2.3) at each sampling site. A total of 136 bands in the visible, near-infrared and short wave infrared regions of the spectrum were analyzed using the leave-one-out cross validation implemented in the Minitab software package. The spectra was divided into different spectral groups e.g. All Bands, Green, Red, NIR, SWIR, All). Individual groups and different group combinations were regressed against each of the soil properties under investigation. The PLS algorithm produces a series of models with different number of factors so that each model in the series contains one additional factor from the previous model. Factors are calculated one at a time, starting with the standardized soil properties and spectral values. The optimal model was identified as the model that produces the highest predicted \( R^2 \) and lowest root mean square error of cross-validation. Once the optimum number of PLS factors had been determined, the model was finalized and the coefficients used to predict spoil properties were identified. The leave-one-out cross validation results, in addition to RPD and \( R^2 \), were used to compare the PLS model results with individual band and simple ratio regression models. These results, including the number of factors and associated bands obtained for the PLS regression model, in addition to model assessment parameters are presented in Table 1 and discussed in the following sections.

3. Results and analysis

In terms of RPD values, the SR models outperformed the NDI models, which in turn outperformed the single-band models (Table 1). The results of the SR and PLS models were mixed. For the POM, sand, labile C, labile N, and stable N, the SR models showed better prediction ability in comparison to the PLS models (Table 1). For the remaining soil properties, PLS model results were superior to the SR models. For the best single band model results, bands 0.59 \( \mu \)m (moisture, bulk density, and MAOM), 0.71 \( \mu \)m (POM, sand, labile C, and labile N), and 0.98 \( \mu \)m (stable C, and stable N) were optimal (Table 1). These bands were identified also in the best NDI and SR models. The optimal predictor combinations of PLS models followed a similar trend, where a few band ranges 0.53–0.59 \( \mu \)m, 1.71–1.79 \( \mu \)m, and 2.20–2.33 \( \mu \)m (for moisture, bulk density, POM and sand), and 0.53–0.59 \( \mu \)m and 1.71–2.09 \( \mu \)m (for labile C and labile N), were identified as predictor variables of multiple soil properties.

The best models of POM, silt + clay, labile C, and labile N have RPD values between 1.43 and 1.94. The \( R^2 \) values of the best PLS models were higher compared to their SR counterparts (Fig. 3). Conversely, the SR models outperformed the PLS models in terms of RMSECV and RPD values for POM, labile C, and labile N. In the case of silt + clay, the \( R^2 \) and RPD were highest and the RMSECV were lowest (better) for the optimal PLS models compared to the other models. The variables, or bands, in the best SR models for labile C and labile N were similar (Fig. 3). Band 0.53 \( \mu \)m is common in these models. The other bands of these models were 2.28 and 2.11 \( \mu \)m respectively, belonging to SWIR part of the spectrum. Band 0.53 \( \mu \)m
is also one of the two best SR variables for the POM. In the silt + clay PLS model, bands 0.59 μm, 1.79 μm, and 0.54–0.56 μm were the most important predictors in the PLS model (Fig. 4). Bands 0.56 and 0.57 μm were the most important bands in PLS models of stable C (Fig. 4). Of all the models tested in the whole study, the SR model of labile N has the highest RPD value (Fig. 3).

For the remaining soil properties the RPD values of the best models are less than 1.4. For moisture, bulk density, sand, stable C, and stable N the RPD values of their respective best models ranged between 1.22 and 1.31 (Table 1). Finally, the results of MAOM are poorest in comparison to others (Table 1), the RPD values of its best models ranged from 0.97 to 1.09. Fig. 5 shows spatial distributions (prediction) maps of the POM, silt + clay, labile C, and labile N soil properties using the regression models having highest RPD values.

4. Discussions

The partitioning of soil mineral and organic matter between fine and coarse particle sizes influences how soils function within ecosystems. The partitioning of soil carbon and nitrogen pools between stable and labile fractions similarly influences soil function. Generally, these forms of partitioning are indicative of nutrient and organic matter turnover and retention, more so than are the sizes of total SOM, carbon, and nitrogen pools. This distinction may be particularly true in coastal wetlands (Middelburg et al., 1996; Morris et al., 2012). Spectral models that can predict these functional soil attributes might therefore be useful for spatially extensive monitoring and management of coastal ecosystems experiencing climate, land use, and sea level change.

Our results demonstrate that in subtropical coastal wetlands, soil properties related to nutrient and organic matter storage and turnover can be modeled successfully with spectral information. However, they also demonstrate that these models vary substantially in predictive accuracy and band composition. The ability to predict soil properties from reflectance data differed among soil variables and among spectral models. For instance, our index of model predictive accuracy (RPD) was 1.94 for labile N but only 1.21 for stable N when considering the best SR model for each soil variable, and the RPD for labile N dropped sharply to 1.05 when considering the best single-band model for that variable. Similarly, the wavelengths that best correlate with soil data also differ among soil variables and among spectral models. For example, the two wavelengths composing the best SR model were 0.53 and 0.70 μm.
for stable C, and were 1.51 and 2.21 μm for stable N, while the best single-band predictor of stable N was 0.98 μm.

Several of the important bands identified in our study are consistent with the results of other spectral models of soil properties. In Bakenhof floodplain, Kooistra et al. (2003) found that the 0.53 and 2.28 μm were important wavelengths in organic matter and clay estimation, and 2.29 μm was significant for organic matter modeling exclusively. Selige et al. (2006) identified the 2.11 and 2.22 μm bands for soil nitrogen from a sample acquired from soilscape undulated terrain that is covered with loess, and coarse sand to fine sand, loamy and clayey sediments from alluvial plain (glacial valley), which are close to the optimal bands (2.11 and 2.21 μm) identified to model soil nitrogen in our study of subtropical intertidal wetlands. The 2.26 and 2.28 μm wavelengths in our labile carbon study are consistent with Henderson et al. (1992), where the 2.26 μm and 2.28–2.29 μm wavelengths were correlated with soil organic matter in agriculture soils. To model substrate soil through vegetation reflectance, Kooistra et al. (2004) found that wavelengths between 0.7 and 1.1 μm were important for organic matter, with bands between 0.7 and 0.8 μm producing the highest regression coefficient. These results are partially consistent with our findings of bands 0.70 and 0.71 as significant bands in modeling POM.
Table 1

<table>
<thead>
<tr>
<th>Soil properties/variables</th>
<th>Best single-band models</th>
<th>Best dual-band models</th>
<th>Best PLS models</th>
<th>Best NII models</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR (%)</td>
<td>0.59</td>
<td>0.25</td>
<td>0.44</td>
<td>0.42</td>
</tr>
<tr>
<td>N (%)</td>
<td>0.59</td>
<td>0.21</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td>POM (mg N/g soil)</td>
<td>0.59</td>
<td>0.26</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>MAOM (%)</td>
<td>0.59</td>
<td>0.34</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>SR + clay (%)</td>
<td>1.79</td>
<td>0.46</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Labile C (mg C/g soil)</td>
<td>0.71</td>
<td>0.29</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Stable C (mg C/g soil)</td>
<td>0.71</td>
<td>0.29</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Labile N (mg N/g soil)</td>
<td>0.71</td>
<td>0.23</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>Stable N (mg N/g soil)</td>
<td>0.71</td>
<td>0.23</td>
<td>0.04</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Because soils are strongly shaped by and influence the properties of plant species, and because satellite remote sensing of areas with dense vegetative cover does not actually capture reflectance directly from soil surfaces, it is important to consider the vegetation properties (biophysical or biochemical) that sensitize the spectral wavelengths or bands correlated with soil properties. Wavelengths 0.53–0.58 μm, 0.63 μm, and 0.70–0.72 μm are sensitive to leaf chlorophyll and chlorophyll a concentrations (Gitelson et al., 1996; Datt, 1999; Gitelson and Merzlyak, 1996; Miller et al., 1990; Vaiphasa et al., 2005). Bands from these ranges are prominent in the best models of POM, potentially indicating the importance of foliage (relative to roots) as a source for this fraction of soil organic matter.

The wavelength 2.11 μm of labile C is an absorption region of the spectra that is attributed to the N–H bonds in amino acids and proteins of the leaf (Martin and Aber, 1997). Band 2.28 μm, of labile nitrogen, is associated with the C–H stretch/CH2 deformation and starch and cellulose (Curran, 1989). There are two possible explanations for why the wavelengths attributed to these foliar biomolecules predict labile soil materials, and both are related to the tight correspondence between ecosystem primary productivity and decomposition (Raich and Schlesinger, 1992; Janssens et al., 2010). First, the presence of these biomolecules in foliage may result from abundant labile soil organic matter. High primary productivity requires rapid protein synthesis in leaves, which in turn depends on a ready supply of soil resources that are best provided by labile organic matter. Second, these biomolecules actually create labile soil material, as foliage with these chemical characteristics (relatively high protein and cellulose, rather than lignin, content) decompose rapidly when it enters the litter (MacLean and Wein, 1978; Gessner and Chauvet, 1994; Chapin et al., 2011). The most important bands of the PLS analysis of stable C are 0.56 and 0.57 μm, which as previously mentioned are attributed to chlorophyll concentration. The most important bands of the PLS model of silt + clay is the 0.59 μm band attributed to leaf pigment absorption as explained above, the 1.79 μm band sensitive to the nitrogen concentration in the leaf (Kalacska and Sanchez-Azofeifa, 2008), and the 0.54–0.56 μm range of bands sensitive to chlorophyll concentration.

We suspect that the statistical models of the soil properties with RPD values less than 1.4 may have resulted from the limited variability in the dependent variables instead of the failure of the Hyperion spectra to predict them. The wavelength region 0.61 μm (correlated with moisture and bulk density) is sensitive to the leaf nitrogen concentration (Read et al., 2002). The 0.98 μm band (correlated with stable C and stable N), and the 1.65 μm band (correlated with MAOM) are sensitive to leaf water content, and may have been indirectly impacted by protein, nitrogen, lignin, cellulose, sugar, and starch of the leaf (Gong et al., 2003; Thenkabail et al., 2004). The wavelength 1.51 μm of stable N is an absorption feature attributed to the foliar protein and nitrogen (Curran, 1989).

The wavelengths of the best models identified from statistical analysis belong to the areas of the spectra that are sensitive to the foliar health and growth. The correlation between the soil properties and these wavelengths reinforce the inference that the reflectance of the canopy and the soil properties are interrelated (Ehrenfeld et al., 2005). Kooistra et al. (2004) performed a study comparable to ours with a similar sample size on SOM and soil moisture but with a different land cover. This research reported R2 values slightly less than ours. The prediction ability index, RPD, of the best models in our study are comparable to other studies by Kooistra et al. (2004) and Piekarczyk et al. (2012) that were focused on bare and substrate soils. Our results thus provide optimism for spectral modeling soils covered by plant canopies. The RPD values of the silt + clay of our study are slightly greater than those found by Piekarczyk et al. (2012). Bare cropping and pasture soil properties...
estimation from Hyperion spectra by Gomez et al. (2008) had R² and RPD results similar to our study. The RPD values of the remaining models are consistent with the studies by Chang et al. (2001) and Summers et al. (2011).

Although we used the Hyperion Spectrometer on board the Earth Observing 1 (EO1) satellite, which is few years beyond its planned lifetime and characterized by low signal to noise imagery, we believe that our results form a basis for analyzing other systems expected for launch in the upcoming several years. Systems such as the Hyperspectral Imager (HIS) sensor of the Environmental Mapping and Analysis Program (EnMAP) system (EnMAP, 2014) and the NASA’s Hyperspectral Infrared Imager or (HyspIRI) mission (HyspIRI, 2014) represent next generation hyperspectral imaging systems expected to provide images with more frequent and better quality imagery that can support soil property modeling and mapping. Additionally these results can assist efforts dedicated to design specialized multispectral imaging system targeting specific applications such as soil property mapping through vegetation cover.

The limited sample size of our study has restricted our ability to perform validation of the models on separate datasets. Instead we substituted it with RMSECV based RPD calculation. The limited sample size is attributed largely to the labor involved in accessing and working at coastal wetland field sites, and in the time-consuming nature of the soil variables that we quantified from in situ samples. Nevertheless, we believe that the results presented in this study showed the feasibility of modeling soil properties using reflectance values of hyperspectral imagery in coastal wetlands leading to the production of soil property maps as shown in Fig. 5. We recognize that the spatial and temporal extent of our study do not suggest global use of the obtained models. However, we believe that our study presents a method and emphasizes the potential to map soil properties from remote sensing techniques using limited field sampling and justify investing in broader scale work with more intensive sampling and spectral analysis of soil properties in tidal coastal environments. The study identifies specific wavelengths and statistical models to use and casts light on the biophysical and biogeochemical processes underlying the creation and use of these models. A larger scale study that utilizes imagery and field samples on a larger scale is recommended as a next step toward regional/global models that can be integrated with the new generation of hyperspectral imaging sensors scheduled for launch in the near future.

5. Conclusions

We tested for the first time the relationship between the reflectance of vegetation acquired from Hyperion (hyperspectral) and Thematic Mapper (multispectral) remote sensing imagery and several soil properties of coastal wetlands environments. In particular, we focused on functional soil properties related to organic matter and nutrient turnover and storage. Single band, two-band index and PLS statistical models were tested. The results showed that the success of Hyperion image models to estimate soil properties based on vegetation reflectance in coastal wetlands are comparable with that of other remote sensing models of bare and vegetation covered soils. The best statistical models were obtained using two-band simple ratio index and partial least-squares analysis. The best Hyperion models for POM, silt + clay, labile C, and labile N, have RPDs between 1.43 and 1.94, and other soil properties had RPDs less than 1.4. Best results were obtained for the SR model of labile N (0.53 µm, 2.11 µm) with R² = 0.81, RMSECV = 0.28, and RPD = 1.94. This ability to predict labile N is notable because labile N is one of the most important soil attributes for ecosystem function. Labile organic N is the principal reservoir for mineralized soil N, which in turn supports the majority of primary production in N-limited coastal wetlands (Nedwell et al., 1994; Anderson et al., 1997; Alongi et al., 2002) and, in fact, in most terrestrial and marine ecosystems (Chapin et al., 2011).

Acknowledgments

We thank Kristine L. Jimenez for technical assistance with field and laboratory work and data management, and thank Joseph M. Latvis for reviewing the article. This research was supported by National Science Foundation grant DEB 1059236.

References


