ISPRS Journal of Photogrammetry and Remote Sensing 148 (2019) 221–234

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Mapping salt marsh soil properties using imaging spectroscopy

Caiyun Zhang[a,*], Deepak R. Mishra[b], Steven C. Pennings[c]

[a] Department of Geosciences, Florida Atlantic University, Boca Raton, FL 33431, USA
[b] Center for Geospatial Research, Department of Geography, University of Georgia, Athens, GA 30602, USA
[c] Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA

ARTICLE INFO

Keywords:
Salt marsh
Soil properties
Imaging spectroscopy
Machine learning
Object-based modeling

ABSTRACT

Tidal salt marshes sequester and store blue carbon at both short and long time scales. Marsh soils shape and maintain the ecosystem by supporting complex biogeochemical reactions, deposition of sediment, and accumulation of organic matter. In this study, we examined the potential of imaging spectroscopy techniques to indirectly quantify and map tidal marsh soil properties at a National Estuarine Research Reserve in Georgia, USA. A framework was developed to combine modern digital image processing techniques for marsh soil mapping, including object-based image analysis (OBIA), machine learning modeling, and ensemble analysis. We also evaluated the efficacy of airborne hyperspectral sensors in estimating marsh soil properties compared to spaceborne multispectral sensors, WorldView-2 and QuickBird. The pros and cons of object-based modeling and mapping were assessed and compared with traditional pixel-based mapping methods. The results showed that the designed framework was effective in quantifying and mapping three marsh soil properties using the composite reflectance from salt marsh environment: soil salinity, soil water content, and soil organic matter content. Multispectral sensors were successful in quantifying soil salinity and soil water content but failed to model soil organic matter. The study also demonstrated the value of minimum noise fraction transformation and ensemble analysis techniques for marsh soil mapping. The results suggest that imaging spectroscopy based modeling is a promising tool to quantify and map marsh soil properties at a local scale, and is a potential alternative to traditional soil data acquisition to support carbon cycle research and the conservation and restoration of tidal marshes.

1. Introduction

Coastal ecosystems such as tidal marshes play an important role in maintaining global biodiversity and human well-being by providing a range of benefits and services (Howard et al., 2014). Numerous studies have highlighted their values in sequestering “blue carbon” which encompasses the carbon within the soil, the living biomass aboveground and belowground, and the non-living biomass (Mcleod et al., 2011). A coastal ecosystem’s capacity to sequester and store carbon is closely related to soil properties which not only impact the short-term scale (years to decades) of carbon sequestration in living plant biomass, but also determine the long-term scale (centuries to millennia) of carbon storage within the sediments. Despite the benefits and services of blue carbon ecosystems, they are threatened due to anthropogenic activities and sea level rise (Mishra et al., 2012; NERRS, 2017). For example, 340,000–980,000 ha of coastal marshes are destroyed each year (Murray et al., 2011). Globally, efforts are underway to restore and create new coastal marshes to mitigate historic and ongoing losses. In U.S., the National Estuarine Research Reserve System (NERRS) has been established with a network of 29 coastal sites designated to protect and study estuarine systems including tidal marshes. Data on soil properties such as soil salinity, soil water content and soil organic matter content are regularly collected and used by NERRS and other entities in wetland conservation and restoration.

Soil salinity is a fundamental parameter controlling and maintaining the structure of tidal marsh communities (Odum, 1988). Increases in the frequency and extent of coastal flooding triggered by sea level rise not only alter soil salinity but also change its spatial distribution which, in turn, affects the marsh distribution, biodiversity, and nutrient uptake of species in marshes (Bradley and Morris, 1991). Soil water content is also an important property affecting the availability of water for the plants, and the direct exchange of soil water with the atmosphere (e.g., evaporation). High evaporation and low water content cause the formation of salt pans devoid of vegetation in coastal marshes (Pennings and Bertness, 2001). Soil organic matter content directly relates to carbon storage in blue carbon ecosystems. Moreover, soil organic

* Corresponding author.
E-mail address: czhang3@fau.edu (C. Zhang).

https://doi.org/10.1016/j.isprsjprs.2019.01.006
Received 3 April 2018; Received in revised form 7 January 2019; Accepted 8 January 2019

matter originating from plant litter decomposition is lighter and bulkier than mineral soils, and has large pore spaces that increase the water holding capacity of the soil. To understand the spatial and temporal dynamics of salt marsh productivity and carbon storage, it is helpful to monitor these parameters across the landscape at regular intervals.

Traditional approaches used in collecting soil data are labor-intensive and time-consuming. In addition, field sampling is inevitably spatially constrained, and samples may not be representative of the true spatial variability of soil properties across a large study domain. Remote sensing could be an alternative approach to understanding soils from field to space. It has been proven effective in modeling and mapping soil properties at a broad scale (Anderson and Croft, 2009; Mulder et al., 2011). Studies have shown that hyperspectral data are more accurate than multispectral data in characterizing soil properties due to their fine spectral resolution (Ben-Dor et al., 2009). Past applications of hyperspectral remote sensing in soil science can be grouped into two categories. The first category is the analysis and modeling of soil properties using direct hyperspectral reflectance of soil samples or bare soils. Point spectroscopy has long been used in the laboratory/field to identify diagnostics spectral features and develop soil models, as reviewed by Nocita et al. (2015). In contrast, application of imaging spectroscopy is limited, and often constrained to homogeneous terrestrial regions or bare soil agricultural fields (e.g., DeTar et al., 2008; Stevens et al., 2008, 2010; Bartholomeus et al., 2011; Bayer et al., 2016; Steinberg et al., 2016; Vaudoura et al., 2016; Žížala et al., 2017). The second category is the application of indirect indicators to retrieve soil properties. Vegetation indices have been linked to terrain and soil types and time series of a vegetation index can indicate the change of soil patterns (Mulder et al., 2011). In the coastal environments, marsh soil properties are well coupled with marsh plant distribution, diversity, and state of health (Odum, 1988), therefore, it is expected that the composite reflectance from marshes which include reflectance from background moist soil and marsh canopy can indirectly indicate soil properties. Zhang et al. (2011) evaluated several vegetation indices of halophyte species derived from a point spectroscopy to estimate marsh soil salinity. Application of imaging spectroscopy for estimating marsh soil properties has not been published. In addition, how much variance of each marsh soil property (e.g., salinity, water content, and organic matter content) can the marsh reflectance explain remains unknown.

Analytical techniques have been used to quantify the relationship between soil properties and remote sensing variables, with parametric regression methods (e.g., partial least square regression) being the most commonly used. A recent study from Forkuor et al. (2017) showed that parametric models were problematic in the prediction of soil properties in un-sampled areas, while non-parametric machine learning models were valuable in mapping regions with sparse or limited samples. There is a need to expand the use of machine learning modeling techniques into soil imaging spectroscopy as an alternative to parametric algorithms. To date, all studies related to mapping soil properties using optical imagery have focused on matching *in-situ* samples with individual pixels to develop soil models, and then generating per-pixel soil maps. In practice, pedologists and ecologists are less interested in a single pixel than a configuration of many pixels that comprise a meaningful patch or region or landscape. Studies also have shown that models developed by matching field samples with regions rather than individual pixels are more robust (Selige et al., 2006). Object-based image analysis (OBIA) is a promising method for such applications. It segments imagery into relatively homogeneous objects for analysis which can reduce the "salt-and-pepper" effect in mapping heterogeneous regions (Blaschke, 2010). Zhang et al. (2018a) indicates that the object-based modeling and mapping of marsh biomass is more effective than the traditional pixel-based methods. OBIA also offers an opportunity to match field samples to a relatively homogeneous region rather an individual pixel. Such an object-based matching scheme has several advantages (Zhang et al., 2018a). First, it can reduce the uncertainty of positional discrepancy between the image and *in-situ* samples. Second, a relatively homogeneous soil patch/object is more representative than any pixel within this patch/object. Third, the local spectral noise/variation can be effectively reduced. Lastly, additional object-based spatial features (e.g., texture) can be extracted for each object, which may have the potential to improve the estimation accuracy of soil properties by including the spatial attributes in the model.

The main objective of this study is to explore the capability of an airborne hyperspectral sensor and simulated spaceborne multispectral sensors to indirectly predict and map soil properties of tidal marshes. Narrowband hyperspectral data might produce an acceptable result for estimating soil properties. But an evaluation of spaceborne multispectral sensors such as WorldView-2 and QuickBird with similar spatial resolution to the airborne hyperspectral imagery is valuable. Such type of satellites allows the monitoring of marsh soil properties at a regular frequency provided there was coincident field data for calibration. Comparison of hyperspectral and multispectral data can help understand the effects of spectral scale in soil mapping. We evaluated contemporary image processing and modeling techniques by combining object-based image analysis, machine learning, and ensemble analysis in a novel framework to effectively map marsh soil properties. We explored four machine learning regression methods and one parametric algorithm, including Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbor (k-NN), and Multiple Linear Regression (MLR). Our logic was that these models might generate similar results but different predictions, thus an ensemble analysis to combine the predictions from each model could produce more robust estimations. The specific objectives of this study were to (1) evaluate whether imaging spectroscopy is more effective in predicting marsh soil properties than multispectral sensors such as WorldView-2 and QuickBird; (2) examine whether machine learning regression algorithms are better than parametric algorithms; (3) assess the benefits of object-based techniques for modeling and mapping soil properties compared with pixel-based methods; and (4) explore the potential of ensemble analysis in mapping tidal marsh soil properties.

2. Study area and data

2.1. Study area

We worked at one of the NERRS, the Sapelo Island in coastal Georgia, USA (Fig. 1). The Reserve encompasses habitats typical of the Carolinian biographic region, which spans the south Atlantic coastline of the United States from North Carolina to upper Florida. The Carolinian region is characterized by expansive tidal salt marshes protected by a chain of barrier islands. Sapelo Island and its surrounding marshes have been the focus of ecological and geological research since the early 1950s. The Duplin River, a 13-km long tidal inlet, flows into Doboy Sound and forms the western boundary of Sapelo Island. The study site is dominated by *Spartina alterniflora* marsh, a habitat containing a wide range of soil conditions typical of salt marshes, but not the low-salinity and higher-organic conditions typical of tidal fresh marshes further upstream in the estuary (Craft, 2007). The site is within the study domain of the Georgia Coastal Ecosystems Long Term Ecological Research Program (https://gce-lter.marsci.uga.edu/) which aims to understand the patterns and processes that shape change in estuarine and marsh environments.

2.2. Data

Data sources include hyperspectral imagery, a marsh species map, and soil data collected by field sampling and laboratory analysis. The hyperspectral data (Hladik, 2013) were collected using the Airborne Imaging Spectrometer for Applications (AISA) Eagle sensor on 20 June 2006 along the Duplin River by the University of Nebraska's Center for Advanced Land Management Information Technologies (CALMIT). The AISA Eagle sensor is a high performance imaging system that

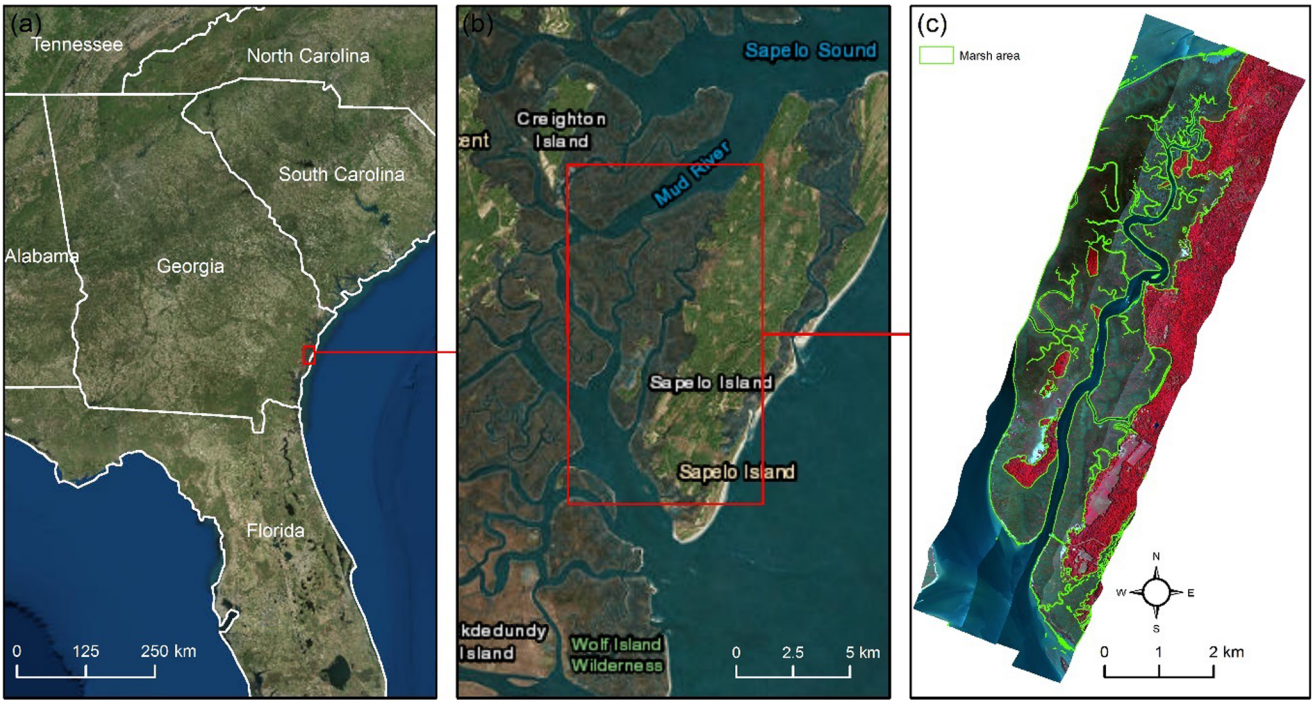

Fig. 1. Location of the study site (a) in the southeast United States, (b) on the coast of Georgia, and (c) along the west side of Sapelo Island. Panel (c) shows a false color composite of the 1-m hyperspectral imagery used in this study, and a marsh mask produced by Hladik et al. (2013).

concurrently collects both visible and near-infrared (VNIR) data with 63 spectral channels within the wavelength range of 400–980 nm. Four parallel flight lines of images at a spatial resolution of 1 m were captured for the study domain. Images were radiometrically corrected using the Fast-Line-sight Atmospheric Analysis of Spectral Hypercube (FLAASH) algorithm, and geometrically corrected using GPS and altitude data. Due to the large data volume, we only mapped a portion of the region highlighted in Fig. 1. We used a marsh map generated by Hladik et al. (2013) to mask out non-marsh areas (e.g., forests, creeks, and rivers). This marsh map was produced using the same hyperspectral imagery as this study and had an overall classification accuracy of 90%. Soil data were collected in the field coincident with the hyperspectral campaign. A total of 346 salt marsh plots (1 × 1 m) were surveyed along 24 transects, and root-zone soil samples were collected over marsh regions for lab analysis. The location of each plot was recorded using a GPS unit. Some samples were collected in the salt pans where high soil salinities exclude vegetation (Pennings and Bertness, 2001). Samples were processed in the laboratory to determine soil water content using the gravimetric method, soil salinity using the rehydration approach, and soil organic matter content using the ignition method (Pennings, 2008). Salinity was calculated as practical salinity units (PSU).

3. Methodology

3.1. A framework for soil modeling and mapping using imaging spectroscopy

We designed a framework to model and map soil properties using an object-based ensemble analysis and imaging spectroscopy (Fig. 2). Hyperspectral data contains redundant spectral information, and the Minimum Noise Fraction (MNF) algorithm is commonly used as a feature selection approach to reduce hyperspectral data. It applies two cascaded principal component analyses, with the first transformation decorrelating and rescaling noise in the data, and the second transformation creating coherent eigenimages that contain useful information, and generating noise-dominated eigenimages (Green et al., 1988). We applied the MNF transformation and selected the first 15 eigenimages based on visual inspection and evaluation of the corresponding eigenvalues, which were then stacked as 15-band MNF imagery to be segmented for object generation and object-based variable extraction. We spatially matched the *in-situ* soil samples to the image objects, leading to a matched dataset for soil model development.

We used four machine learning algorithms, ANN, SVM, RF, and k-NN, and one parametric method, MLR, to develop the models, and the Correlation Coefficient (r), Mean Absolute Error (MAE), Percent Mean Absolute error (PMAE), and Root Mean Squared Error (RMSE) to assess the performance of each model. We conducted model comparisons and significance tests to determine: (a) the need of ensemble analysis; and (b) which models to include in the ensemble analysis. We used each model to predict soil properties for all marsh objects using spectral features and ensemble analysis to combine the outputs from individual models. We also used the ensemble analysis to generate an uncertainty map to spatially delineate the overall uncertainty caused by modeling methods.

To compare the efficacy of hyperspectral sensors with spaceborne multispectral sensors, we simulated the 8-band WorldView-2 and 4-band QuickBird data from the 63-band AISA hyperspectral imagery based on the sensor-specific spectral filter functions. We spatially matched the simulated WorldView-2 and QuickBird data to the *in-situ* samples to develop all aforementioned soil models. To compare the object-based modeling with the pixel-based modeling approach, we also matched the *in-situ* soil samples with the corresponding pixels for pixel-based modeling. The key steps in the modeling framework included image segmentation, data matching between *in-situ* samples and hyperspectral imagery, model development and soil mapping, and model evaluation and accuracy assessment. These steps are detailed in the following subsections.

3.2. Image segmentation

We applied the multi-resolution segmentation algorithm to the 15-band MNF imagery in eCognition Developer 9.0 (Benz et al., 2004; Trimble, 2014) to generate image objects. The multi-resolution segmentation approach needs to set a scale parameter, color/shape, and

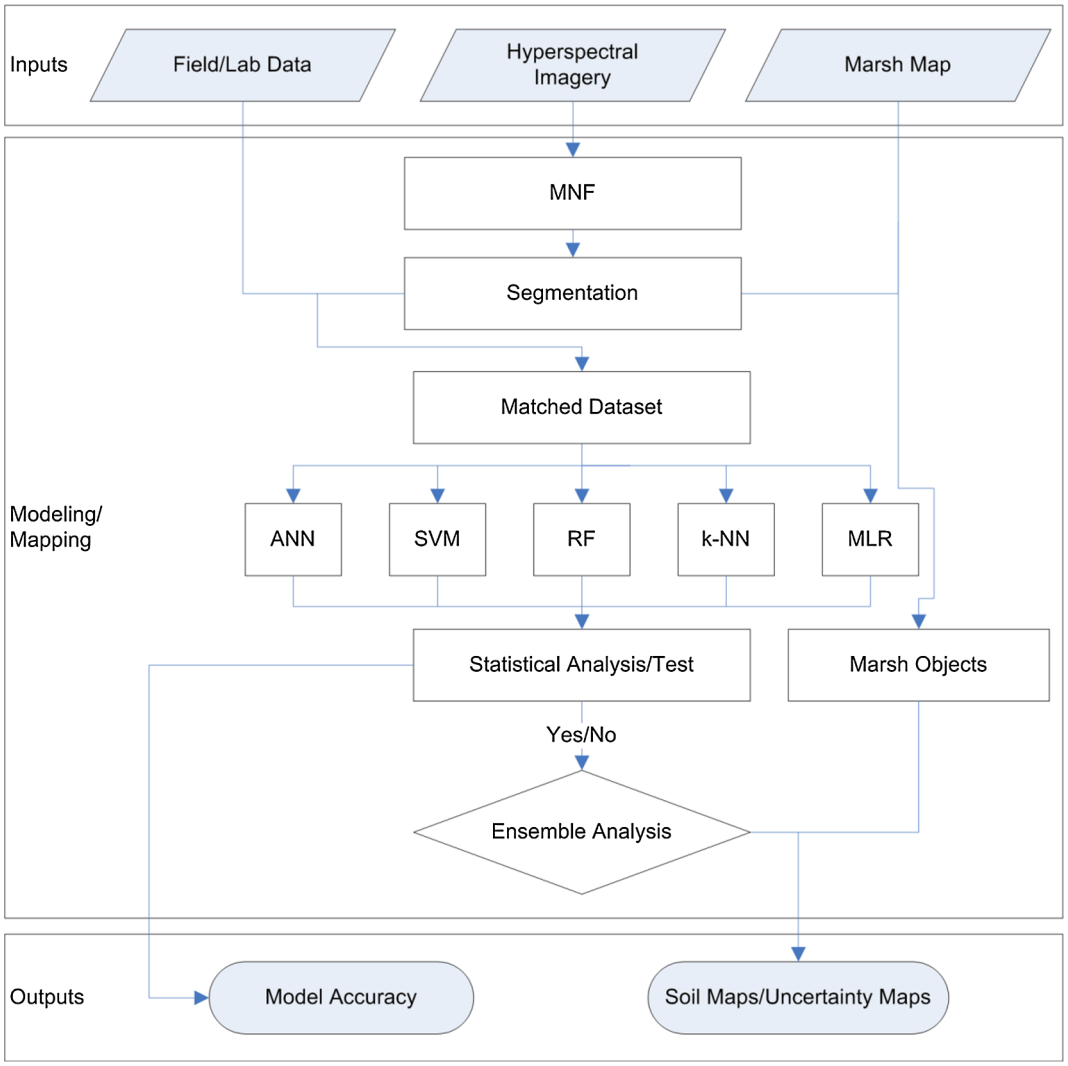

Fig. 2. Designed framework for modeling and mapping marsh soil properties using hyperspectral data and object-based ensemble analysis.

smoothness/compactness weights, among which the scale parameter is the most important variable. Several approaches have been developed to optimize the scale parameter in the segmentation (e.g., Drăguţ et al., 2010; Johnson and Xie, 2011; Grybas et al., 2017). In this study, we applied the approach developed by Johnson and Xie (2011) to determine an optimal scale parameter. This approach begins with a series of segmentations using different scale parameters, and then identifies the optimal image segmentation using a method that takes into account global intra-segment and inter-segment heterogeneity measures. We found that a scale of 15 was optimal for the study site in the segmentation. The weight of each MNF layer was set based on its eigenvalue produced in the MNF transformation. Color/shape weights were set to 0.9/0.1 so that spectral information would be considered most heavily for segmentation. Smoothness/compactness weights were set to equal weights. Following the segmentations, we extracted the spectral features of each object which were then used for modeling and mapping.

3.3. Data matching between in-situ samples and hyperspectral imagery

We matched the *in-situ* samples to an image object to create a spatially coincident dataset for object-based model development. For a comparison purpose, we also matched the *in-situ* samples with individual pixels of the original 63-band AISA hyperspectral imagery, 15-band MNF transformed imagery, 8-band WorldView-2 imagery, and 4-band QuickBird imagery, respectively, to develop pixel-based models.

3.4. Model development and soil mapping

We compared four machine learning regression algorithms, ANN, SVM, RF, and k-NN, and one parametric regression method, MLR, for predicting soil attributes. ANN is an important technique in machine learning and modeling. Various ANN algorithms have been developed and applied in remote sensing (Mas and Flores, 2008). In this study, we used the multilayer perceptron algorithm of ANN, a commonly used approach in remote sensing. SVM is a statistical learning approach (Vapnik, 1995) which has been widely used in remote sensing (Mountrakis et al., 2011). RF is a decision tree based ensemble approach which constructs numerous small regression trees contributing to the predictions. Breiman (2001) provided a detailed description of RF and Belgiu and Drăguţ (2016) reviewed the application of RF techniques in remote sensing. k-NN is a relatively simple approach. The estimation is predicted as a weighted average value with k spectrally nearest neighbors using a weighting method (Chirici et al., 2016). MLR has been frequently used in predicting soil attributes. It assumes that remotely sensed independent variables are linearly related to *in-situ* samples. Each algorithm requires the specification of several parameters and the setting of these parameters will impact the results. In this study, each algorithm was implemented and tuned in Waikato Environment for Knowledge Analysis (WEKA), a machine learning modeling software package (Hall et al., 2009). The experimenter function in WEKA can determine the best model for any algorithm with

different parameter specifications based on several statistical metrics such as correlation coefficient. We tuned each algorithm based on this function to select the best model for our datasets.

To integrate the predictions from different models for soil mapping, we used a weighted combining scheme developed by Zhang et al. (2018a) based on the correlation coefficient (r) derived from each model. If model i ($i = 1, 2, \ldots M$) has a correlation coefficient r_i, then the final prediction P is calculated as:

$$P = \sum \frac{r_i p_i}{\sum_{i=1}^{M} r_i} \quad (1)$$

where p_i is the prediction of model i, and M is the total number of models in the ensemble analysis. In this way, a model with a larger r obtains a higher weight, and the sum of weights is 1.0. If a negative estimation occurs from a model, the prediction from this model will be eliminated by setting a weight of zero in the ensemble analysis.

3.5. Model evaluation and accuracy assessment

We used the k-fold cross validation technique for model training and testing. This evaluation method has proven valuable in machine learning techniques (Anguita et al., 2012). It splits the sampling data into k subsets first, and then iteratively, one subset (1, 2, … k) is used to assess the model and other remaining subsets are used to train the model. In this study, k was specified as 10, a number commonly used in the literature to divide the samples into 10 divisions. In the iteration, a subset was excluded from training the model (i.e. 9 subsets were used as training data) to be used to generate the estimation of this subset. Each model would iterate 10 times. After the iteration, soil predictions were generated for all *in-situ* locations, which could be then used to calculate the statistical metrics, including r, MAE, PMAE, and RMSE. They were calculated as:

$$r = \frac{\sum_{i=1}^{N} (p_i - \bar{p_i})(p_{lab} - \bar{p_{lab}})}{\sqrt{\sum_{i=1}^{N} (p_i - \bar{p_i})^2} \sqrt{\sum_{i=1}^{N} (p_{lab} - \bar{p_{lab}})^2}} \quad (2)$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |p_i - p_{lab}| \quad (3)$$

$$PMAE = \frac{\frac{1}{N} \sum_{i=1}^{N} |p_i - p_{lab}|}{\bar{p_{lab}}} \times 100 \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (p_i - p_{lab})^2} \quad (5)$$

where p_i is the model prediction; p_{lab} is the lab soil value; $\bar{p_i}$ and $\bar{p_{lab}}$ are the mean of model predictions and mean of lab soil values, respectively. N is the total number of matched samples.

In the k-fold cross validation, each model would produce the predictions for all the corresponding *in-situ* samples. The outputs from each model and lab soil data could be then used to determine which model should be included in the ensemble analysis using ANOVA F-test. Following the strategy in Zhang et al. (2018a), we built a full linear model first with the models as the independent variables and lab soil data as the dependent variable, and then compared it with a reduced model which was nested within the full model. We built a series of full and reduced models and compared them using the ANOVA F-test to determine which models could be combined in the ensemble analysis.

If ensemble analysis of two models is applied for the final estimation of soil properties, an uncertainty map can be produced by mapping the absolute bias between these two models. But if there are three or more models to be combined, more uncertainty maps will be produced, which might not be informative. We developed a process to produce only one uncertainty map by considering the predictions from all the models in the ensemble analysis. The process started by classifying the ensemble estimations into several categories (e.g., ten categories for each soil property in this study) to be used for mapping each property, and then calculated the total votes of each category from the models for an input image object. For example, if all five models are combined in ensemble analysis, then one vote means no agreement is obtained for an input object for estimating its soil properties, i.e. all five models produced completely different estimations for this input image object, suggesting the prediction of soil properties for this image object is challenging. If the total number of votes is 2, 3, or 4, it means 2, 3, or 4 models produced similar estimations for an input object. If the total number of votes is 5, it means a full agreement is achieved, i.e., all five models produced similar estimations for the input image object, suggesting the estimation for this object is the most reliable.

4. Results

4.1. Marsh soil properties from laboratory analyses

The Duplin River tidal salt marsh sediments consist mainly of reworked Pleistocene muds, eroded and redeposited by the tidal currents (Chalmers, 1997). The sediments are characterized by high and variable salinities, water content, and organic matter content, with a mean salinity of 44 PSU, water content of 50%, and organic matter content of 14% (Table 1). Soil salinity varied from 6 to 306 PSU, with the highest values in salt pans. Soil water and organic matter content were also highly variable with ranges of 10–75% and 1–41%, respectively. Soil water and organic matter content were strongly correlated ($R^2 = 0.8$) with an exponential fit, and weakly related to soil salinity (Fig. 3). The relationships were slightly different when we used samples from vegetated marshes only (i.e., mud salt pans excluded), but the results still showed that soil water and organic matter content were strongly related to each other and weakly related to soil salinity (results not shown). Correlation analysis between each property and the original 63 spectral bands showed that visible region was more related to soil salinity, while NIR was more related to water and organic matter content (Fig. 4). Band 24 with a wavelength of 615.05 nm was found to have the highest correlation coefficient ($r = 0.88$) with salinity, while band 54 with a wavelength of 942.3 nm had the least correlation ($r = 0.54$) with soil salinity. Soil water content and soil organic matter were negatively related to salt marsh reflectance. Band 58 with a wavelength of 932.95 nm was found to have the highest correlation with a r of 0.74 and 0.55 to water and organic matter, respectively. Band 31 with a wavelength of 680.5 nm had the lowest correlation with a r of 0.45 and 0.34 to water and organic matter, respectively.

4.2. Pixel- and object-based modeling using different datasets and algorithms

The pixel-based modeling results from 63-band original AISA hyperspectral data, simulated 8-band WorldView-2 data and 4-band QuickBird data, and 15-band MNF transformed data, and object-based modeling results using the 15-band MNF transformed dataset for predicting soil salinity, water content, and organic matter content are provided in Tables 2–4, respectively. For the soil salinity estimation, application of the original AISA hyperspectral dataset from five models produced r of 0.87–0.92, a MAE of 9.65–12.10 PSU, a PMAE of

Table 1
Descriptive statistics of soil attributes from laboratory analyses.

	Salinity	Water content (%)	Organic matter (%)
Minimum	6	10	1.2
Maximum	306	75	41
Mean	44	50	14
Standard deviation	37	19	9

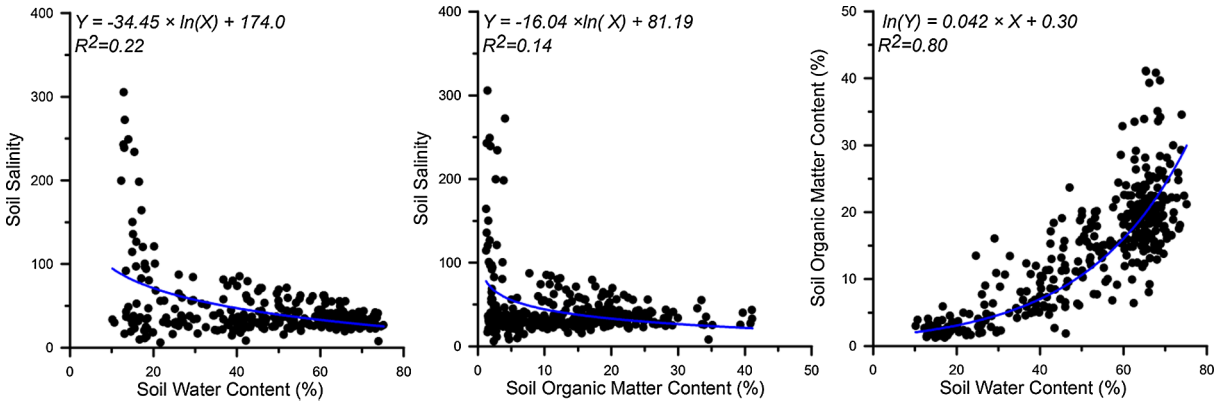

Fig. 3. Bivariate relationships between the three marsh soil properties derived from field sampling. Regression equations and R^2 values are given inside each panel.

21.93–27.50%, and a RMSE of 14.61–18.39 PSU (Table 2). When we resampled the AISA hyperspectral imagery to WorldView-2 satellite data, the performance of prediction was slightly reduced with a *r*, MAE, PMAE, and RMSE of 0.86–0.90, 10.47–12.21 PSU, 23.8–27.75%, and 16.16–18.99 PSU. The performance further deteriorated when the AISA imagery was resampled to QuickBird data with a *r*, MAE, PMAE, and RMSE of 0.86–0.88, 11.20–12.92 PSU, 25.45–29.36%, and 18.03–18.73 PSU, respectively. When the MNF transformed dataset was used, the performance of prediction improved slightly (*r*: 0.90–0.93; MAE: 8.61–11.3 PSU; PMAE: 20.91–25.68%, and RMSE: 14.11–17.31 PSU) compared to the application of original AISA hyperspectral data. Comparison of five algorithms showed that SVM model consistently produced the best result for mapping soil properties using hyperspectral, WorldView-2 and QuickBird datasets when *r*, MAE, PMAE, and RMSE observed jointly, while RF applied to MNF transformed data generated the best predictions of soil salinity. When pixel-based models were compared with the object-based model, we found that the object-based modeling using the 15-band MNF transformed dataset produced the best results with the highest *r* of 0.94, the lowest MAE, PMAE and RMSE of 8.45 PSU, 19.20% and 13.43 PSU. Among the five modeling algorithms, the SVM model produced the highest *r* and k-NN produced the lowest MAE, while RF generated the lowest PMAE and ANN produced the lowest RMSE.

Table 3 shows the modeling results for soil water content estimation. The pixel-based estimation results illustrated that the SVM consistently produced the best result among five algorithms when the original hyperspectral, WorldView-2 and QuickBird datasets were applied. This is similar to the soil salinity modeling. The 63-band hyperspectral imagery produced a good result with a *r*, a MAE, a PMAE and a RMSE of 0.76–0.82, 8.09–10.09%, 16.18–20.18% and 10.99–12.38%, respectively. Application of WorldView-2 dataset was also encouraging with the *r*, MAE, PMAE, and RMSE of 0.65–0.80, 8.56–12.09%, 17.12–24.18% and 11.54–14.45%. Again, the performance of prediction was reduced (*r*: 0.61–0.73; MAE: 10.30–12.39%; PMAE: 20.6–24.78%; and RMSE: 13.35%–15.10%) when QuickBird dataset was used. Application of the MNF transformed dataset at the pixel-level

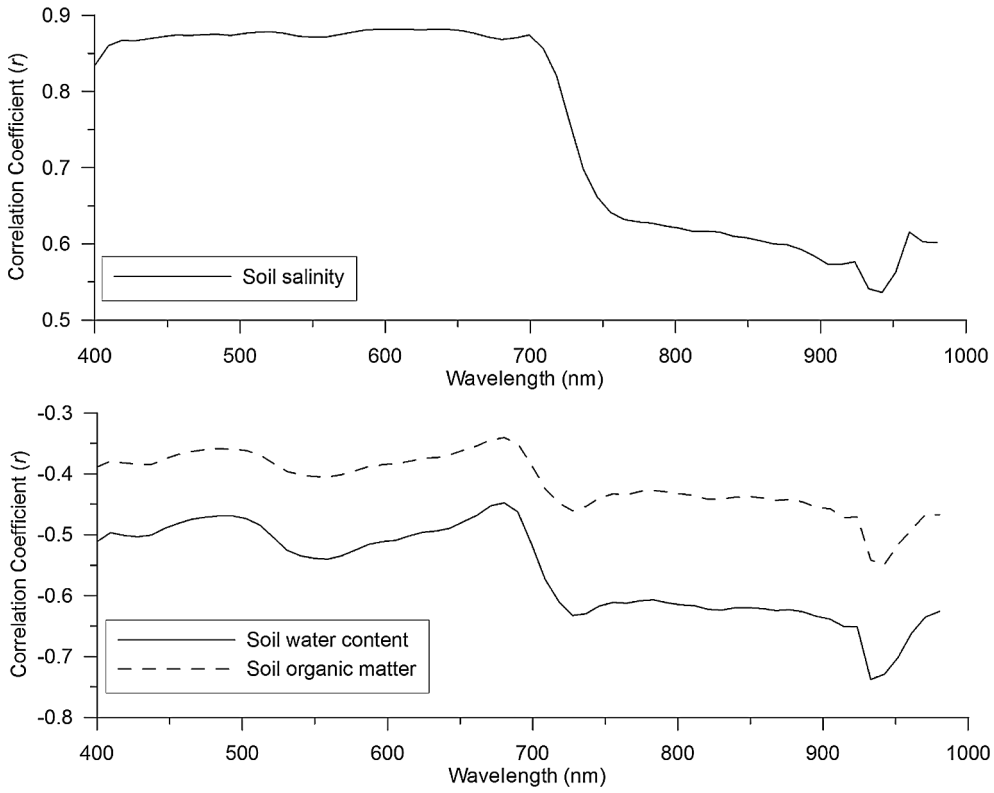

Fig. 4. Correlation analysis between each soil property and reflectance of salt marshes derived from 63-band hyperspectral data.

Table 2
Model performance for soil salinity estimation (unit: PSU) using different datasets and algorithms. The best result of the five algorithms is bolded.

Statistical Metrics	ANN	SVM	RF	k-NN	MLR
Pixel-based modeling					
63-band AISA hyperspectral data					
CC (*r*)	0.90	**0.92**	0.87	0.88	0.89
MAE (PSU)	11.51	**9.65**	11.10	11.41	12.10
PMAE (%)	26.16	**21.93**	25.23	25.93	27.50
RMSE (PSU)	16.21	**14.61**	18.39	17.63	16.78
8-band WorldView-2 multispectral data					
CC (*r*)	0.90	**0.90**	0.86	0.87	0.89
MAE (PSU)	11.50	**10.47**	11.42	11.55	12.21
PMAE (%)	26.14	**23.80**	25.95	26.25	27.75
RMSE (PSU)	16.41	**16.16**	18.99	18.73	17.27
4-band QuickBird multispectral data					
CC (*r*)	0.86	**0.88**	0.87	0.87	0.87
MAE (PSU)	11.81	**11.20**	11.73	11.65	12.92
PMAE (%)	26.84	**25.45**	26.66	26.48	29.36
RMSE (PSU)	18.15	**18.03**	18.68	18.40	18.73
15-band MNF transformed data					
CC (*r*)	0.91	0.91	0.92	**0.93**	0.90
MAE (PSU)	11.05	10.81	9.20	**8.61**	11.30
PMAE (%)	25.11	24.57	20.91	**19.57**	25.68
RMSE (PSU)	15.68	17.31	14.44	**14.11**	16.25
Object-based modeling					
15-band MNF transformed data					
CC (*r*)	0.93	0.94	0.92	0.93	0.93
MAE (PSU)	8.98	8.96	8.53	8.45	9.13
PMAE (%)	20.41	20.36	19.39	19.20	20.75
RMSE (PSU)	13.43	13.71	14.35	14.14	13.83

CC (*r*): Correlation Coefficient; MAE: Mean Absolute Error; RMSE: Root Mean Squared Error; PMAE: Percent Mean Absolute Error.
ANN: Artificial Neural Network; SVM: SVM; RF: Random Forest; k-NN: k-Nearest Neighbor; MLR: Multiple Linear Regression. MNF: MNF.

Table 3
Model performance for soil water content estimation (unit: %) using different datasets and algorithms. The best result of the five algorithms is bolded.

Statistical Metrics	ANN	SVM	RF	k-NN	MLR
Pixel-based modeling					
63-band AISA hyperspectral data					
CC (*r*)	0.80	**0.82**	0.79	0.76	0.76
MAE (%)	8.83	**8.09**	8.83	9.54	10.09
PMAE (%)	17.66	**16.18**	17.66	19.08	20.18
RMSE (%)	11.38	**10.99**	11.56	12.38	12.37
8-band WorldView-2 multispectral data					
CC (*r*)	0.72	**0.80**	0.73	0.71	0.65
MAE (%)	10.61	**8.56**	9.75	10.13	12.09
PMAE (%)	21.22	**17.12**	19.5	20.26	24.18
RMSE (%)	13.36	**11.54**	12.98	13.49	14.45
4-band QuickBird multispectral data					
CC (*r*)	0.61	**0.73**	0.70	0.65	0.63
MAE (%)	12.19	**10.30**	10.27	11.02	12.39
PMAE (%)	24.38	**20.60**	20.54	22.04	24.78
RMSE (%)	15.10	**13.35**	13.63	14.82	14.79
15-band MNF transformed data					
CC (*r*)	0.83	0.86	0.82	0.85	0.82
MAE (%)	8.59	7.31	8.25	7.19	8.67
PMAE (%)	17.18	14.62	16.5	14.38	17.34
RMSE (%)	10.74	9.97	10.87	10.01	10.82
Object-based modeling					
15-band MNF transformed data					
CC (*r*)	0.88	**0.89**	0.88	0.89	0.87
MAE (%)	6.79	**6.08**	6.30	6.51	6.89
PMAE (%)	13.58	**12.16**	12.6	13.02	13.78
RMSE (%)	9.19	**8.67**	8.91	8.86	9.21

Notes: CC (*r*), MAE, RMSE, PMAE, ANN, SVM, RF, k-NN, MLR and MNF are same abbreviations as in Table 2.

improved the prediction of soil water content with the highest *r* of 0.86, the lowest MAE, PMAE, and RMSE of 7.19%, 14.38%, and 9.97%. Object-based modeling increased the *r* to 0.89, and decreased the MAE, PMAE and RMSE to 6.08%, 12.16% and 8.67%, respectively. Comparison between pixel-based and object-based modeling results for water content estimations, again, demonstrated the superior performance of object-based modeling using MNF transformed data when the same regression algorithm was applied.

Table 4 displays the modeling results for soil organic matter estimations. For the pixel-based estimation, the *r*, MAE, PMAE and RMSE were in the range of 0.54–0.61, 5.02–5.58%, 35.86–39.86% and 6.86–7.37%, respectively, using the hyperspectral imagery; and within 0.50–0.61, 5.13–6.01%, 36.64–42.93%, and 6.92%–7.77% using the WorldView-2 dataset. The worst estimation was from the QuickBird data with *r*, MAE, PMAE and RMSE in the range of 0.40–0.52, 5.56–6.42%, 39.71–45.86% and 7.46–8.22%, respectively. MNF transformed data produced the best estimation among the pixel-based modeling but object-based modeling enhanced the estimation with a *r* of 0.80, MAE, PMAE and RMSE of 3.66%, 26.14% and 5.00%, respectively, using the SVM algorithm.

4.3. Object-based mapping of soil properties and uncertainty

The above comparative results showed that object-based modeling using the MNF transformed dataset produced the best result in estimating all three soil properties. ANOVA *F*-test consistently indicated that three algorithms, SVM, RF, and k-NN should be included in the ensemble analysis. Statistical metrics showed that the other two algorithms, ANN and MLR sometimes produced similar results compared to the three chosen algorithms but not consistently. We thus estimated and mapped soil properties using object-based ensemble analysis on the MNF transformed dataset, as shown in Fig. 5(a)–(c) with blue indicating low estimations and red representing high estimations. The three marsh soil properties revealed interesting spatial patterns. In the soil salinity map, a small area was observed to have the highest salinity (more than 60 PSU) in the marsh west of Duplin River. A ridge with intermediate salinity (yellow[1]) occurred parallel to the east bank of the river, while low salinity (blue) was observed over a large portion of the east part of the marsh. Water content and organic matter were well coupled with high values observed in the east part of the river, and low values in the west marsh.

The uncertainty maps produced by ensemble analysis are shown in Fig. 5(d)–(f). The classified categories shown in the corresponding maps reveal the voting agreement between the three models, SVM, RF, and k-NN. If there was only one vote, then no agreement was obtained (in red). If there were two votes for an input object, a partial agreement was produced (in blue). If there were three votes, a full agreement was achieved (in green). The uncertainty maps revealed the consistency and differences in the estimations from three modeling algorithms. In general, the estimation of soil properties was robust with a major portion shown in blue and green, indicating a high confidence in the estimation. A few areas displayed a “warning sign” in red, where no estimation agreement was obtained. These regions had the highest probability of being incorrectly estimated. Comparison of three uncertainty maps showed that the most agreement was produced for soil water content estimation with a larger area in green, whereas, the most disagreement occurred in soil organic matter estimation with a large number of areas in red.

[1] For interpretation of color in Fig. 5, the reader is referred to the web version of this article.

Table 4
Model performance for soil organic matter content estimation (unit: %) using different datasets and algorithms. The best result of the five algorithms is bolded.

Statistical Metrics	ANN	SVM	RF	k-NN	MLR
Pixel-based modeling					
63-band AISA hyperspectral data					
CC (*r*)	0.57	**0.61**	0.57	0.54	0.55
MAE (%)	5.47	**5.02**	5.06	5.53	5.58
PMAE (%)	39.07	**35.86**	36.14	39.50	39.86
RMSE (%)	7.18	**6.86**	7.13	7.37	7.23
8-band WorldView-2 multispectral data					
CC (*r*)	0.51	**0.61**	0.55	0.52	0.50
MAE (%)	5.90	**5.13**	5.29	5.99	6.01
PMAE (%)	42.14	**36.64**	37.79	42.79	42.93
RMSE (%)	7.62	**6.92**	7.31	7.77	7.58
4-band QuickBird multispectral data					
CC (*r*)	0.40	0.52	0.52	0.45	0.46
MAE (%)	6.42	5.69	5.56	6.41	6.24
PMAE (%)	45.86	40.64	39.71	45.79	44.57
RMSE (%)	8.09	7.46	7.54	8.22	7.76
15-band MNF transformed data					
CC (*r*)	0.65	0.68	0.65	**0.69**	0.67
MAE (%)	5.07	4.74	4.73	**4.57**	4.99
PMAE (%)	36.21	33.86	33.79	**32.64**	35.64
RMSE (%)	6.63	6.36	6.61	**6.31**	6.53
Object-based modeling					
15-band MNF transformed data					
CC (*r*)	0.76	**0.80**	0.78	0.79	0.78
MAE (%)	4.06	**3.66**	3.76	3.66	3.95
PMAE (%)	29.00	**26.14**	26.86	26.14	28.21
RMSE (%)	5.47	**5.00**	5.21	5.24	5.35

Notes: CC (*r*), MAE, RMSE, PMAE, ANN, SVM, RF, k-NN, MLR and MNF are same abbreviations as in Table 2.

5. Discussion

5.1. Predicting marsh soil properties using imaging spectroscopy

5.1.1. Marsh soil salinity estimation

Previous studies have found that plant composition can be used as an indirect indicator of soil salinity (Metternicht and Zinck, 2003). Several vegetation indices from hyperspectral data have been examined to quantify soil salinity with varying degree of accuracies (e.g., Zhang et al., 2011). Our study demonstrated that soil salinity can be indirectly predicted in a marsh environment using imaging spectroscopy, machine learning algorithms, and *in-situ* data. Application of the hyperspectral data with the MNF transformation achieved a *r* of 0.94 in soil salinity estimation using the object-based modeling method. The soil salinity map clearly mirrored the plant zonation pattern, which was expected because first, salinity plays an important role in mediating the plant zonation pattern (Pennings et al., 2005), and second species reflectance played a heavy role in the model input. Typically, in a southeastern USA salt marsh, the highest soil salinity is associated with the salt pans and the lowest with *Juncus roemerianus* (Pennings and Bertness, 2001). Soil salinity for short and medium *Spartina alterniflora* and for *Borrichia frutescens* is usually higher than for tall *Spartina alterniflora* because of less frequent tidal flushing that allows evapotranspiration to concentrate salts in the soil. Finally, salt pans occur where soil salinities exceed levels that the plants can tolerate. To examine the relationship between soil salinity and plant species distribution, the species distribution map was overlaid on the hyperspectral color composite and the salinity map covering three representative marsh areas including salt pan, short *Spartina alterniflora,* and tall *Spartina alterniflora*, are highlighted in Fig. 6. In general, the pattern of estimated salinity had a good alignment with the marsh plant distribution. The highest estimated salinity was observed for salt pans, and the lowest salinity for tall *Spartina alterniflora*. Short *Spartina alterniflora* was mainly distributed over the eastern side of the river where medium salinity ridges were observed. The plant distribution map only delineated the boundary of each species, while the salinity map provided more details within each plant species. These details indicate the heterogeneity of soil conditions within an area occupied by a single species. And two may provide new insights into other soil characteristics such as geomorphology, and plant or invertebrate ecology.

5.1.2. Marsh soil water content estimation

Soil water content is another important factor affecting plant zonation patterns in low-latitude coastal marshes (Pennings et al., 2005). We found that marsh soil water content was negatively correlated to soil salinity, as reported in a previous study (Zhang et al., 1997). Using imaging spectroscopy of marsh reflectance to indirectly estimate and map soil moisture has not been explored. In this study, the MNF transformed dataset produced a *r* of 0.89 for predicting soil water content using object-based modeling. Soil water content in the surface sediment varies over time daily with tides, and monthly with the spring-neap tide cycle. As a result, accurate estimates of water content will only be possible given calibrations based on field data for a given time period. There are issues to quantify the absolute water content for a specific time based on vegetation reflectance only. Also, there is time lag between vegetation response and average water content change. However, the indirect method developed here can at least indicate the spatial pattern of average soil water content or the availability of water to sustain marsh plants. The plant species map was overlaid on the soil water content map with two regions (salt pans and medium *Spartina alterniflora*) highlighted in two separate maps. Salt pans had a low water content, which was expected due to the high evaporation and concentration of salt (Fig. 7a). Regions with high water content were distributed along the eastern river bank covered by medium *Spartina alterniflora*. The western bank, however, did not show a parallel pattern. It had moderate soil water content and was covered by *Borrichia frutescens* and short *Spartina alterniflora*. This illustrates that soil water content is not simply a function of distance to the river or tidal inlet, but also of other factors such as elevation.

5.1.3. Marsh soil organic matter estimation

Compared with soil salinity and water content, more efforts have been made to quantify soil organic matter using imaging spectroscopy through the direct methods, and a varying degree of accuracies have been reported with *r* ranging from 0.69 (DeTar et al., 2008) to 0.95 (Selige et al., 2006). In this study, a moderate accuracy was obtained in predicting organic matter content with an *r* of 0.8 using the MNF transformed data and object-based modeling. Compared with the estimation of salinity and water content, a relatively lower accuracy was produced for organic matter estimation. Again, the plant species map was overlaid on the soil organic matter map with salt pans and short *Spartina alterniflora* highlighted. As expected, salt pans had the lowest organic matter content (Fig. 7c). High organic matter was observed over regions dominated by medium and short *Spartina alterniflora* (Fig. 7d). Soil organic matter and water content were well coupled with high values observed on the eastern side of the river dominated by *Spartina alterniflora*, and relatively lower values on the western side of the river dominated by *Salicornia virginica* (Fig. 7). Inconsistency between soil water content and organic matter was also detected. For example, over the eastern side of the river, high water content was observed in the region dominated by medium *Spartina alterniflora* (Fig. 7b), while the high organic matter zone showed a trend of eastward migration from medium *Spartina alterniflora* to short *Spartina alterniflora*. Soil organic matter is characterized by slow temporal dynamics, and as such is far more stable than soil salinity and water content. Therefore, maps of soil organic content will only need to be produced every few years.

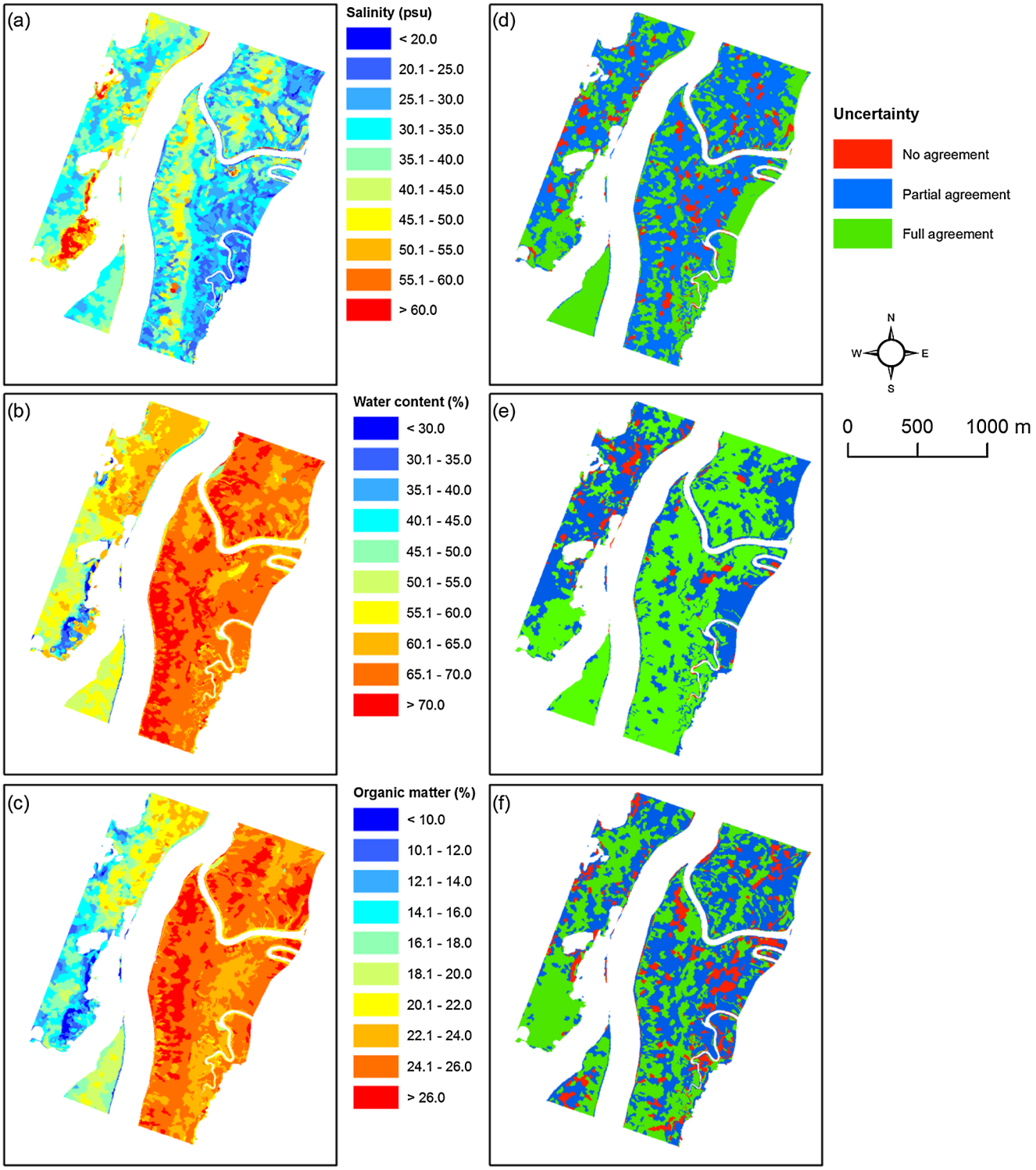

Fig. 5. Maps of estimated soil salinity (a), soil water content (b), and soil organic matter content (c), and corresponding uncertainty maps generated from ensemble analysis of the predictions from SVM, RF, and k-NN models for estimating soil salinity (d), soil water content (e), and soil organic matter content (f).

5.2. Performance of hyperspectral vs. multispectral sensors

Hyperspectral data is available mostly from airborne sensors, and thus data acquisition is expensive and infrequent. It is valuable to assess the capability of spaceborne sensors for higher-frequency monitoring since these sensors could obtain data less expensive and visit the same location regularly. We explored the potential applicability of 8-band WorldView-2 and 4-band QuickBird spaceborne multispectral sensors for modeling and monitoring marsh soil properties. Feature selection techniques are frequently used in imaging spectroscopy to reduce the dimensionality of hyperspectral data. Few studies have evaluated the impact of dimensionality reduction on the prediction of soil properties. We compared 63-band AISA hyperspectral dataset and a simulated 8-band WorldView-2 data, 4-band QuickBird data, and 15-band MNF transformed data to assess the impact of the dimensionality reduction of hyperspectral data in the prediction of soil properties. We found that

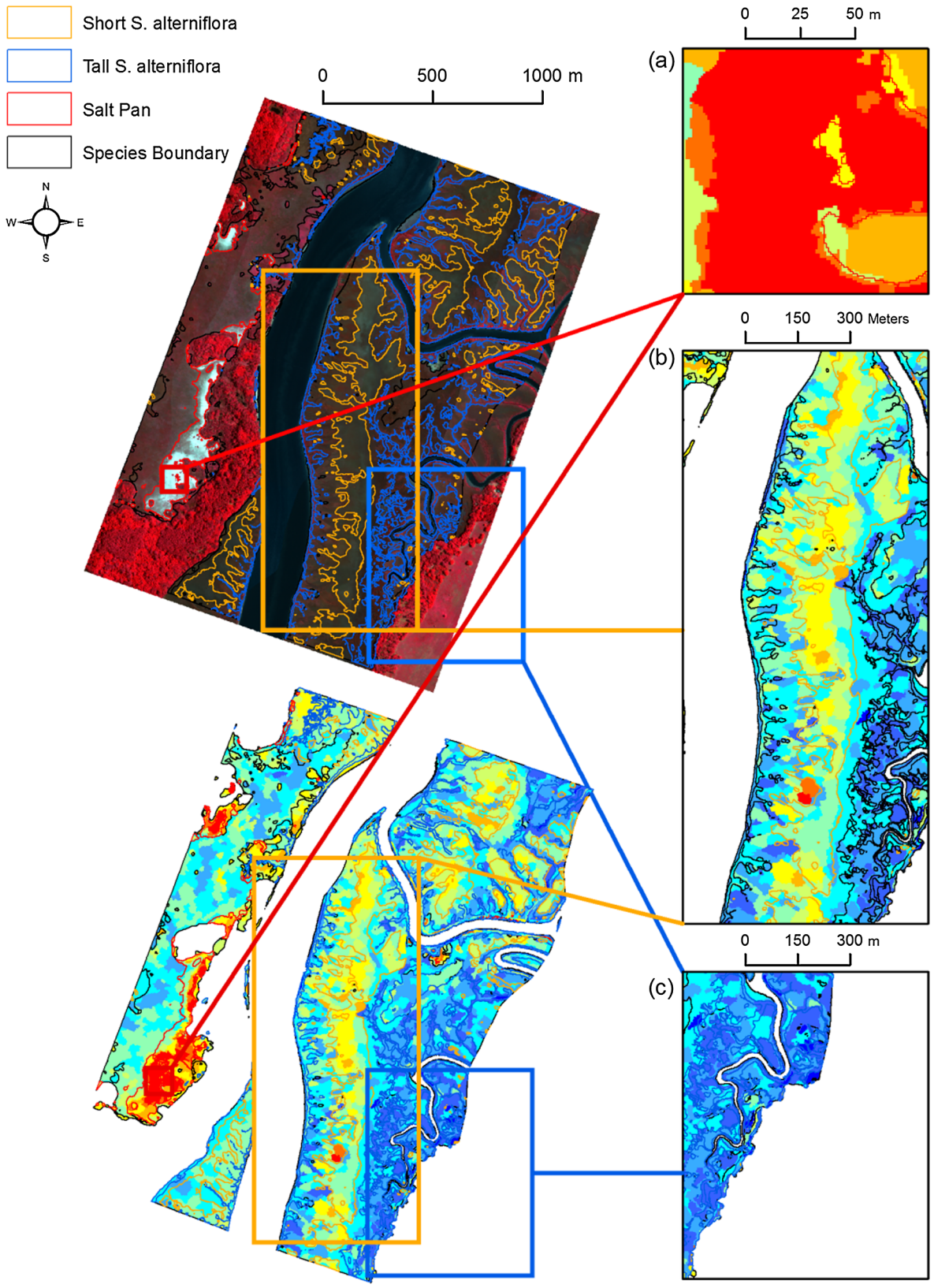

Fig. 6. Zoomed-in areas that highlight salt pans with the highest salinity in patches (a), short *Spartina alterniflora* with higher salinity as salinity ridges (b), and tall *Spartina alterniflora* with low salinity (c). Areas covered by salt pan, short *Spartina alterniflora*, tall *Spartina alterniflora* and other species are derived from the classification of hyperspectral imagery (Hladik et al., 2013). Color scales are same as Fig. 5 for soil property mapping.

estimation accuracy was somewhat declined when the spectral resolution was reduced. Although a slightly reduced accuracy was obtained from WorldView-2 and QuickBird datasets compared with the hyperspectral data, encouraging accuracies (r: 0.90 and 0.88) were produced for soil salinity estimation. For soil water content estimation, the results were also acceptable from the two spaceborne sensors with

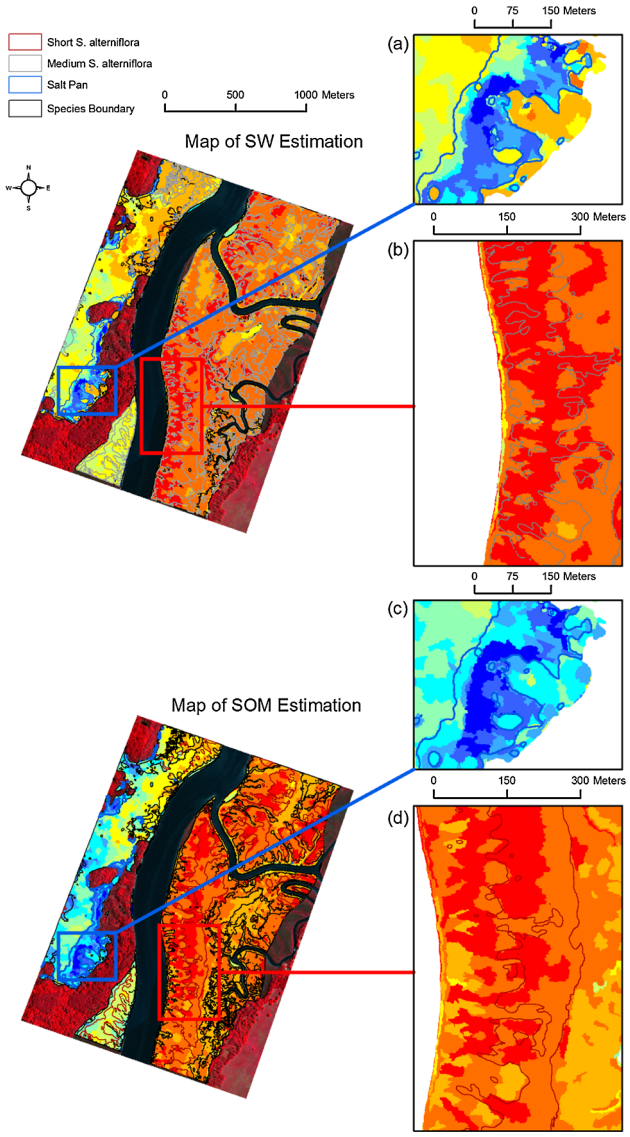

Fig. 7. Zoomed-in areas to highlight salt pans with the lowest water content (a), medium *Spartina alterniflora with* high water content (b) along the river bank, salt pans with the lowest organic matter content (c), and short and medium *Spartina alterniflora* with high organic matter content (d). Color scales are same as Fig. 5 for soil property mapping.

an *r* of 0.80 and 0.73, respectively. But poor results were produced for organic matter content prediction. Given the high cost in acquiring hyperspectral data, it would be much more cost-effective to use spaceborne sensors such as Worldview-2 and QuickBird for broad area mapping and monitoring of marsh soil properties. These sensors have a high spatial resolution and the multi-temporal nature, which provides the opportunity to monitor the soil properties on any date for which it is convenient to collect field data for calibration. Application of Landsat-type sensors (e.g., Landsat and Sentinel-2) is also possible because it has a similar spectral configuration as QuickBird. But Landsat-type data have a relatively coarse spatial resolution of 30 m and the mixed pixel problem might be a challenge for the modeling and mapping because marsh vegetation, salinity, water content and soil organic matter content can change rapidly over a few meters. The scalability of the soil remote sensing models to true multispectral satellite data and sensitivity of the models to subtle changes in reflectance are important avenues for future studies.

Our study showed that the application of a noise reduction algorithm such as the MNF transformation was important in indirect soil characterization using hyperspectral imagery. Ben-Dor et al. (2002) applied this technique to separate the noise components of the DAIS-7915 data in order to reconstruct the informative spectral channels for soil modeling, rather than to reduce the high dimensionality of hyperspectral data. Research from Belluco et al. (2006) for mapping salt marsh vegetation using hyperspectral data demonstrated that MNF technique was the most effective method to reduce the redundancy in hyperspectral data for marsh identification. It extracts information-containing bands and identifies the noise component in the general case of unequal noise content over different bands, which makes it more effective than other feature selection techniques such as principle component analysis. This technique not only helps by reducing the dimensionality of hyperspectral data to save computational cost, but also largely improves the modeling and mapping accuracy. In addition, if object-based modeling and mapping is applied, segmenting the original hyperspectral imagery will be computation-intensive. Application of MNF transformed imagery will largely reduce the computational cost in the segmentation. Similarly, a range of feature selection techniques and indices have been developed to reduce the dimensionality of hyperspectral data and identify important spectral narrowband regions for modeling or isolating soil characteristics. It would be valuable to examine the effectiveness of those techniques such as red-edge on indirect marsh soil modeling in future research.

5.3. Parametric regression, machine learning, and ensemble analysis

Past studies commonly applied the partial least square regression and MLR methods in soil modeling. Stevens et al. (2010) were the first to explore the SVM regression in soil organic carbon mapping using imaging spectroscopy. We used four machine learning models including ANN, SVM, RF, and k-NN, and compared the results with the MLR method. In general, SVM had the best performance in processing each dataset, while other methods produced comparable results in accuracy. All models achieved acceptable accuracies for estimating soil salinity and produced low accuracies for estimating organic matter content when the original hyperspectral, and a simulated WorldView-2, and QuickBird datasets were used. Although all models produced similar results in terms of the statistical metrics based on the validation data, the predictions from each model for mapping were different. For example, for salinity estimation, the difference between two models was more than 80 for some regions. These types of discrepancies were also observed for the same objects in predictive models of water content and organic matter content. ANN and MLR even produced results similar as the machine learning algorithms. However, when they were applied for mapping, many outliers were produced including negative predictions, especially over the regions with sparse samples. This was another reason why we did not include ANN and MLR in the ensemble mapping. The diversity in predictions among models was a result of differences in the architecture of the algorithms. Each model has its pros and cons, and an ensemble analysis can make the prediction more robust than the application of an individual model alone, especially for unknown regions (Zhang et al., 2018a). We explored the potential of ensemble analysis for estimating soil properties. It is unnecessary to recalculate *r*, MAE and RMSE from the ensemble analysis, because these statistical metrics should be similar to the models that are combined. The ensemble analysis also provides a solution to map the spatial uncertainty to complement the traditional metrics in soil modeling. The uncertainty map from ensemble analysis effectively identifies regions that are easy to estimate and other areas that are difficult to quantify. This type of map is useful to guide field soil sampling. For example, more field samples should be collected from the regions with a high uncertainty to recalibrate or tune the models. Use of uncertainty maps generated from ensemble analysis is somewhat limited and further research is needed to investigate the potential application of such maps in soil modeling and

mapping. This problem was also reported in Forkuor et al. (2017), and that is another reason we did not include it in the ensemble analysis.

5.4. Pros and cons of object-based modeling and mapping

Our study is the first attempt to examine the potential of object-based modeling for mapping tidal marsh soil properties. Object-based modeling produced a higher accuracy than the pixel-based method and generated informative soil maps, which suggests that it is a promising alternative to the traditional pixel-based modeling and mapping. It improved the estimation accuracy of all three soil properties. This is attributed to two major factors. First, our object-based approach spatially matched field samples to an image object, and thereby reduced the positional discrepancy to combine field data with remote sensing imagery. The misregistration of field samples can strongly affect the estimation accuracy, especially when a fine spatial resolution image is used. Both GPS readings of the field locations and georeferencing of the imagery have errors, particularly at the edges of a pixel, leading to the positional discrepancy between two datasets. There is a higher probability that a field sample is located in an object/patch rather than within a single pixel. Thus, using an object is more robust than a pixel and can potentially increase the estimation accuracy. Second, as demonstrated in previous OBIA studies on wetlands, the object-based modeling approach can reduce local noise and heterogeneity, and thereby enhance the accuracy of the analysis (Dronova, 2015). Using the mean spectrum of an object can reduce the effect of noise, thus increasing the accuracy of estimations. Object-based modeling has also been attempted over bare soil regions using imagine spectroscopy and those studies have recommended ways to reduce the noise issue in the pixel-based modeling (Ben-Dor et al., 2002; Selige et al., 2006). To the best of our knowledge, object-based soil mapping has not been conducted.

There are other potentials using OBIA in soil modeling. The object-based approach offers the spatial features (e.g., texture) which have proven valuable for marsh species classification (e.g., Zhang et al., 2018b). The object-based spatial features might be valuable for soil modeling too, especially when plants are used as the indirect indicator of soil properties. Another advantage of the object-based method is that it allows for segmenting the image into objects/segments at multiple scales, which offers an opportunity to map soil properties at multiple scales or at different levels of detail adaptable to various applications. In contrast, the pixels are uni-scale and represent a fixed area on the ground with a single value. It would be a valuable direction for future research to include object-based spatial features and multi-scale mapping features to map salt marsh soil properties. There are some limitations to the use of object-based modeling such as the parameter specification in the image segmentation. The settings of these parameters will impact the results. A range of methods have been developed to optimize the scale parameter. But other parameters are commonly set empirically or subjectively. It is worthy to evaluate the impact of segmentation on the modeling and mapping result.

5.5. Sources of errors

We have identified multiple potential sources of errors in our soil modeling and mapping procedure, including upscaling, time gaps between acquisition of field samples and hyperspectral data, radiometric mismatch between flight lines, and limited field samples. First, we matched the field samples collected from a homogeneous plot (1 m × 1 m) to the image objects with varying shapes and sizes. Advantages of this matching scheme have been discussed in the above subsection, but uncertainties from the segmentation would affect the modeling and mapping results. Second, there was a short time gap between the acquisition of hyperspectral imagery (6/20/2006) and field samples (6/20/2006–6/26/2006). This should not be an issue for organic matter estimation but could have influenced the salinity and water content results because these two attributes are temporally more dynamic. Third, there was a slight radiometric mismatch of hyperspectral data between flight lines. Such radiometric differences are often caused by varied atmospheric condition, sun angles, and illumination differences, rather than by physical changes in surface properties (Jensen, 2015). Although we applied atmospheric correction to each flight line, the mismatch could still be observed over the seam line when the flight lines were mosaicked. Several relative radiometric correction methods have proven to be useful to remove this mismatch from multispectral imagery, but none of these techniques generated ideal results for the hyperspectral data used in this study. This mismatch could have impacted both the modeling and mapping results particularly around the seamline. In addition, due to the large volume of the hyperspectral data, the images were divided into small tiles. Inconsistencies in mapping were also observed along seam lines between tiles. This is a typical issue when using a large volume hyperspectral dataset in soil mapping. The fourth and last error could be from the limited number of field samples. *In-situ* sampling and laboratory analysis is time-consuming and labor-intensive. In this study, field samples were collected along transects to represent the varying structure of salt marsh soils. However, it was difficult to collect samples over regions that were hard to access on the ground. Estimations over those regions should be reevaluated in future with more field data. Despite the issues that could have affected modeling accuracy, the general spatial patterns of the three soil properties in the study domain seem appropriate from a geomorphological and ecological prospective.

5.6. Implications for marsh ecology and carbon cycling studies

This study has broader implications in ecology and carbon cycle studies pertaining to tidal marsh Gross Primary Production (GPP) modeling. GPP models developed based on a method proposed by Monteith (1972) have been parametrized widely by ecologists for tidal wetlands by incorporating variables such as soil salinity, soil water content (a proxy for tidal fluctuations), temperature, and phenology (Yuan et al., 2007; Kathilankal et al., 2008; Moffett et al., 2010; Barr et al., 2013; Forbrich and Giblin, 2015; Knox et al., 2018). Although these types of parametrizations have been found to work well within the footprint of an Eddy Covariance flux tower deployed in a tidal wetland, they may not be ideal for scaling up a GPP model to an area outside of the footprint. That is because the tidal fluctuations and salinity data used in upscaling light use efficiency models are based on few discrete stationary measurements and may not represent the conditions of the tower footprint. The approach presented in this study would improve the satellite based GPP models because it includes the per-pixel/object values of soil salinity and water content rather than assuming that these parameters remain the same throughout the entire scene. Whether or not the soil salinity, water and organic matter content products can enhance GPP models and improve the estimation of other biophysical parameters such as aboveground and belowground biomass remains to be examined. It would also be useful to assess the sensitivity of the developed models to changes in canopy reflectance caused by plant phenology.

6. Conclusions and future research

We developed an alternative framework to traditional interpolation techniques for estimating marsh soil properties by using simultaneously collected hyperspectral imagery and limited *in-situ* data. We combined object-based image analysis (OBIA), machine learning, and ensemble analysis techniques in the framework. We evaluated the potential applicability of the spaceborne multispectral sensors WorldView-2 and QuickBird for modeling and monitoring temporal variations of marsh soil properties. An acceptable modeling result was achieved, and the spatial patterns of marsh soil salinity, water content, and soil organic matter were revealed. The results illustrated that imaging spectroscopy

is valuable for soil data collection. Spaceborne WorldView-2 and QuickBird sensors look promising in monitoring marsh soil salinity and water content. The logic of the framework can be used to other marsh ecosystems to delineate marsh soil patterns.

Direct application of the developed models across time might be problematic due to the high temporal variability of soil salinity and water content, low variation of soil organic matter, and the time lag of vegetation response to the change of these properties. More research is needed to address the suitable recalibration time interval for these models, which depend on the temporal variability of the soil and plant properties within a study site. For example, soil organic matter changes slowly (years) compared to soil salinity and soil water content (based on daily and spring-neap tidal cycles) and plant reflectance patterns (based on phenology; seasonally). Therefore, the "train once and classify ever" type of framework typically used in time-series analysis is not suitable for this study. Additional work is needed to develop an optimal time interval (weekly, monthly or seasonally) to train the models with independent field data and coincident hyperspectral/multispectral data in order to capture the spatial and temporal variability of these soil parameters. It might also be informative to include more modeling algorithms in the ensemble analysis. The inclusion of other datasets such as topography and environmental variables might also increase the prediction performance. We hope this study will stimulate the application of modern remote sensing data processing techniques to soil spectroscopy in general, and particularly marsh soil mapping.

Acknowledgements

This material relied upon spectral and field data whose collection was supported by the National Science Foundation, USA, through the Georgia Coastal Ecosystems Long-Term Ecological Research program under Grant Nos. OCE-OCE-9982133, OCE-0620959 and OCE-1237140. We thank the reviewers for their helpful comments on the manuscript. This is contribution number 1063 from the University of Georgia Marine Institute, USA.

References

Anderson, K., Croft, H., 2009. Remote sensing of soil surface properties. Prog. Phys. Geogr. 33, 457–473.
Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., Ridella, S., 2012. The 'K' in K-fold cross validation. ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, April 25–27.
Bartholomeus, H., Kooistra, L., Stevens, A., Leeuwen, M., Wesemael, B., Ben-Dor, E., Tychon, B., 2011. Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. Int. J. Appl. Earth Observ. Geoinform. 13, 81–88.
Barr, J.G., Engel, V., Fuentes, J.D., Fuller, D.O., Kwon, H., 2013. Modeling light use efficiency in a subtropical mangrove forest equipped with CO_2 eddy covariance. Biogeosciences 10, 2145–2158.
Bayer, A.D., Bachmann, M., Rogge, D., Müller, A., Kaufmann, H., 2016. Combining field and imaging spectroscopy to map soil organic carbon in a semiarid environment. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 9, 3997–4010.
Belgiu, M., Drăguţ, L., 2016. Random Forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.
Belluco, E., Camuffo, M., Ferrari, S., Modenese, L., Silvestri, S., Marani, A., et al., 2006. Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing. Rem. Sens. Environ. 105, 54–67.
Ben-Dor, E., Patkin, K., Banin, A., Karnieli, A., 2002. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data. A case study over clayey soils in Israel. Int. J. Rem. Sens. 23, 1043–1062.
Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., Sommer, S., 2009. Using imaging spectroscopy to study soil properties. Rem. Sens. Environ. 113, S38–S55.
Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multiresolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J. Photogramm. Rem. Sens. 58, 239–258.
Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Rem. Sens. 65, 2–16.
Bradley, P.M., Morris, J.T., 1991. The influence of salinity on the kinetics of NH4 + uptake in *Spartina alterniflora*. Oecology 85, 375–380.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Chalmers, A.G., 1997. The Ecology of the Sapelo Island National Estuarine Research Reserve. National Oceanic and Atmospheric Administration, Office of Coastal Resource Management, Sanctuaries, and Reserves Division.
Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E.O., Waser, L.T., et al., 2016. A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. Rem. Sens. Environ. 176, 282–294.
Craft, C., 2007. Freshwater input structures soil properties, vertical accretion, and nutrient accumulation of Georgia and U.S. tidal marshes. Limnol. Oceanogr. 52, 1220–1230.
DeTar, W.R., Chesson, J.H., Penner, J.V., Ojala, J.C., 2008. Detection of soil properties with airborne hyperspectral measurements of bare fields. Trans. ASABE 51, 463–470.
Drăguţ, L., Tiede, D., Levick, S.R., 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. Int. J. Geograph. Inform. Sci. 24, 859–871.
Dronova, I., 2015. Object-based image analysis in wetland research: a review. Rem. Sens. 7, 6380–6413.
Forbrich, I., Giblin, A.E., 2015. Marsh-atmosphere CO2 exchange in a New England salt marsh. Biogeosciences 120, 1825–1838.
Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. PLoS ONE 12, e0170478.
Green, A.A., Berman, M., Switzer, P., Craig, M.D., 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Trans. Geosci. Remote Sens. 26, 65–74.
Grybas, H., Melendy, L., Congalton, R.G., 2017. A comparison of unsupervised segmentation parameter optimization approaches using moderate- and high-resolution imagery. GISci. Rem. Sens. 54, 515–533.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. SIGKDD Explor. 11, 10–18.
Hladik, C., 2013. AISA Hyperspectral Imagery of the GCE Domain for Vegetation. Georgia Coastal Ecosystems LTER Project, University of Georgia, Long Term Ecological Research Network. https://doi.org/10.6073/pasta/09075b66ccf91a0c87ffbd113f754977.
Hladik, C., Schalles, J., Alber, M., 2013. Salt marsh elevation and habitat mapping using hyperspectral and LIDAR data. Rem. Sens. Environ. 139, 318–330.
Howard, J., Hoyt, S., Isensee, K., Pidgeon, E., Telszewski, M., 2014. Coastal Blue Carbon: Methods for Assessing Carbon Stocks and Emissions Factors in Mangroves, Tidal Salt Marshes, and Seagrass Meadows. Conservation International, Intergovernmental Oceanographic Commission of UNESCO, International Union for Conservation of Nature, Arlington, Virginia, USA.
Jensen, J.R., 2015. Introductory Digital Image Processing: A Remote Sensing Perspective, fourth ed. Prentice Hall, Upper Saddle River, NJ.
Johnson, B., Xie, Z., 2011. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. ISPRS J. Photogramm. Rem. Sens. 66, 473–483.
Kathilankal, J.C., Mozdzer, T.J., Fuentes, J.D., D'Odorico, P., McGlathery, K.J., Zieman, J.C., 2008. Tidal influences on carbon assimilation by a salt marsh. Environ. Res. Lett. 3, 044010.
Knox, S.H., Windham-Myers, L., Anderson, F., Sturtevant, C., Bergamaschi, B., 2018. Direct and indirect effects of tides on ecosystem-scale CO2 exchange in a brackish tidal marsh in Northern California. Biogeosciences 123.
Mas, J.F., Flores, J.J., 2008. The application of artificial neural networks to the analysis of remotely sensed data. Int. J. Remote Sens. 29, 617–663.
Metternicht, G.I., Zinck, J.A., 2003. Remote sensing of soil salinity: potentials and constraints. Remote Sens. Environ. 85, 1–20.
Mcleod, E., Chmura, G.L., Bouillon, S., Salm, R., Björk, M., Duarte, C.M., et al., 2011. A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering CO_2. Front. Ecol. Environ. 9, 552–560.
Mishra, D.R., Cho, H.J., Ghosh, S., Fox, A., Downs, C., Merani, P.B.T., et al., 2012. Postspill state of the marsh: remote estimation of the ecological impact of the Gulf of Mexico oil spill on Louisiana salt marshes. Remote Sens. Environ. 118, 176–185.
Moffett, K.B., Wolf, A., Berry, J.A., Gorelick, S.M., 2010. Salt marsh–atmosphere exchange of energy, water vapor, and carbon dioxide: effects of tidal flooding and biophysical controls. Water Resour. Res. 46, W10525.
Monteith, J.L., 1972. Solar radiation and productivity in tropical ecosystems. J. Appl. Ecol. 9, 747–766.
Mountrakis, G., Im, J., Ogole, C., 2011. SVMs in remote sensing: a review. ISPRS J. Photogramm. Rem. Sens. 66, 247–259.
Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping – a review. Geoderma 162, 1–19.
Murray, B.C., Pendleton, L., Sifleet, S., 2011. State of the Science on Coastal Blue Carbon: A Summary for Policy Makers. Institute for Environmental Policy Solutions Report NIR 11-06, pp. 1–43.
NERRS, National Estuarine Research Reserve System. < https://coast.noaa.gov/nerrs/ > (last access on August 24, 2018).
Nocita, M., Stevens, A., Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., et al., 2015. Soil spectroscopy: an alternative to wet chemistry for soil monitoring. Adv. Agron. 132, 139–159.
Odum, W.E., 1988. Comparative ecology of tidal freshwater and salt marshes. Annu. Rev. Ecol. Syst. 19, 147–176.
Pennings, S.C., Bertness, M.D., 2001. Salt marsh communities. In: Bertness, M.D., Gaines, S.D., Hay, M.E. (Eds.), Marine Community Ecology. Sinauer Associates, Inc., Sunderland, MA. USA, pp. 550.
Pennings, S.C., Grant, M., Bertness, M.D., 2005. Plant zonation in low-latitude salt marshes: disentangling the roles of flooding, salinity and competition. J. Ecol. 93, 159–167.
Pennings, S.C., 2008. Vegetation and Invertebrate Communities in 500 Plots in the Duplin

and Dean Creek watersheds: Ground Truth Data for Matching Hyperspectral Imagery. Georgia Coastal Ecosystems LTER Project, University of Georgia, Long Term Ecological Research Network. https://doi.org/10.6073/pasta/640c9b83f76147c1afec56bf9a0252fc.

Selige, T., Boehner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma 136, 235–244.

Steinberg, A., Chabrillat, S., Stevens, A., Segl, K., Foerster, S., 2016. Prediction of common surface soil properties based on Vis-NIR airborne and simulated EnMAP imaging spectroscopy data: prediction accuracy and influence of spatial resolution. Rem. Sens. 8, 613.

Stevens, A., Wesemael, B., Btholomeus, H., Rosillon, D., Tychon, B., Ben-Dor, E., 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. Geoderma 144, 395–404.

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., Wesemael, B., 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. Geoderma 158, 32–45.

Trimble, 2014. eCognition Developer 9.0.1 Reference Book.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Vaudoura, E., Gilliota, J.M., Belb, L., Lefevrec, J., Chehdica, K., 2016. Regional prediction of soil organic carbon content over temperate croplands using visible near-infrared airborne hyperspectral imagery and synchronous field spectra. Int. J. Appl. Earth Observ. Geoinform. 49, 24–38.

Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L.L., Baldocchi, D., et al., 2007. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. Agric. For. Meteorol. 143, 189–207.

Zhang, C., Denka, S., Cooper, H., Mishra, D.R., 2018a. Quantification of sawgrass marsh aboveground biomass in the coastal Everglades using object-based ensemble analysis and Landsat data. Rem. Sens. Environ. 204, 366–379.

Zhang, C., Denka, S., Mishra, D.R., 2018b. Mapping freshwater marsh species in the wetlands of Lake Okeechobee using very high-resolution aerial photography and lidar data. Int. J. Rem. Sens. 39, 5600–5618.

Zhang, T.T., Zeng, S., Gao, Y., Ouyang, Z., Li, B., Fang, C., Zhao, B., 2011. Using hyperspectral vegetation indices as a proxy to monitor soil salinity. Ecol. Ind. 11, 1552–1562.

Zhang, M., Ustin, S.L., Rejmankova, E., Sanderson, E.W., 1997. Monitoring Pacific coast salt marshes using remote sensing. Ecol. Appl. 7, 1039–1053.

Žížala, D., Zádorová, T., Kapička, J., 2017. Assessment of soil degradation by erosion based on analysis of soil properties using aerial hyperspectral images and ancillary data, Czech Republic. Rem. Sens. 9, 28.