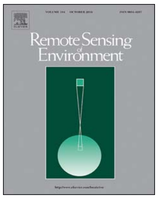
Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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



Quantification of sawgrass marsh aboveground biomass in the coastal Everglades using object-based ensemble analysis and Landsat data

Caiyun Zhang[a,*], Sara Denka[a], Hannah Cooper[a], Deepak R. Mishra[b]

[a] *Department of Geosciences, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA*
[b] *Center for Geospatial Research, Department of Geography, University of Georgia, 210 Field Street, Athens, GA 30602, USA*

ARTICLE INFO

Keywords:
Object-based biomass modeling
Ensemble analysis for biomass prediction
Machine learning regression algorithms
Coastal marshes

ABSTRACT

Large-scale biomass quantification of sawgrass (*Cladium jamaicense*) marsh is critical to understand the carbon and energy cycle in the Florida Everglades. There is also a need to monitor biomass changes in the coastal Everglades due to continuing sea level rise. Previous research in biomass estimation of coastal marshes has focused on pixel-based parametric modeling methods. In this study, an object-based ensemble analysis approach was developed to map sawgrass biomass at multiple scales using Landsat data. Four machine learning regression algorithms including Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbor (k-NN), and Artificial Neural Network (ANN) were evaluated and compared to the commonly used Multiple Linear Regression (MLR) method for both live and total sawgrass biomass estimation. A weighted combining scheme was developed to integrate predictions from comparable models for ensemble analysis. Nonparametric machine learning models had better performance than the parametric approach. ANN and SVM produced similar results in live biomass estimation with the correlation coefficient (r) larger than 0.9, while ANN achieved the best result for the total biomass estimation ($r = 0.94$). Sawgrass biomass maps were produced for two harvest seasons in 2014 and 2016 at three detail levels, which successfully revealed the spatial and temporal (seasonal and interannual) sawgrass biomass variations. Ensemble analysis of the ANN and SVM predictions of live sawgrass biomass not only made the estimation more reliable, but also generated an uncertainty map to identify the regions with a robust biomass prediction, as well as challenging areas for biomass quantification. It is concluded that the object-based ensemble analysis is a promising alternative to the commonly used pixel-based biomass modeling techniques.

1. Introduction

1.1. Significance of sawgrass biomass mapping in the coastal Everglades

The Florida Everglades is the largest subtropical wetland in the USA, supporting many threatened and endangered species. The Comprehensive Everglades Restoration Plan (CERP) was enacted in 2000 to restore, preserve, and protect the Everglades, which has been severely modified by human activities in the past century (Everglades Restoration, 2017). CERP is the largest hydrologic restoration project in the USA with an estimated cost of more than $10.5 billion and a 35 + year timeline. A range of projects have been conducted for Everglades restoration, many of which require plant biomass information, especially for the sawgrass marsh which occupies about 70% of the Everglades (Loveless, 1959). Sawgrass (*Cladium jamaicense*), a tall and perennial sedge in the family of Cyperaceae, is the icon of the Everglades. Estimation of sawgrass biomass not only benefits the understanding of the carbon and energy cycles in the Everglades, but also provides information on plant health condition and standing fuel load for fire management (Lauck and Benscoter, 2015). The Everglades is a peatland with a significant carbon storage coming from on-site plant production. The emergent plant marshes such as sawgrass are particularly productive. Biomass changes have been considered as an indicator of vegetation stress induced by nature and human-caused disturbances (Klemas, 2013). There is a need to monitor the changes in the biochemical and biophysical properties of the Everglades ecosystem, including the plant biomass. Some sawgrass communities may be considered as fire climaxes because they have been burned annually during periods of drought to improve habitat for wildlife and reduce large accumulations of fuels (Loveless, 1959). For the coastal Everglades, this task is more urgent due to sea level rise which might lead to loss of vegetation and gains in low marsh and mudflats (Byrd et al., 2014).

---

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

http://dx.doi.org/10.1016/j.rse.2017.10.018
Received 7 May 2017; Received in revised form 3 September 2017; Accepted 13 October 2017

Please cite this article as: Zhang, C., Remote Sensing of Environment (2017), http://dx.doi.org/10.1016/j.rse.2017.10.018

1.2. Past efforts in remote sensing of biomass for wetland marshes

Current sawgrass biomass data are mainly collected in the field using destructive or non-destructive methods (Childers et al., 2006; Lauck and Benscoter, 2015). This in-situ data collection procedure is time-consuming and labor-intensive, thus limited biomass data have been acquired over a few plots in the coastal Everglades. With the advance of remote sensing techniques, it is anticipated that this procedure can be superseded by digital techniques and plant biomass can be quantified at a large scale. Efforts have been made to estimate and map coastal plant biomass via remote sensing, as reviewed by Klemas (2013). Research in biomass estimation of non-woody marshes can be grouped into two categories. The first is the application of hand-held radiometer/spectroradiometer data (e.g., Hardisky et al., 1983, 1984; Zhang et al., 1997; Kearney et al., 2009; Trilla et al., 2013). The main purpose of this type of work is to identify the optimal spectral channels or indices for marsh biomass prediction. The developed models can hardly be used for large-scale biomass mapping. The second direction is the application of airborne/spaceborne sensors with the main objective to model and map biomass (Gross et al., 1987; Zhang et al., 1997; Jensen et al., 1998, 2002; Byrd et al., 2014; Kulawardhana et al., 2014; Ghosh et al., 2016; Kim et al., 2016; Lumbierres et al., 2017). Studies have demonstrated that the airborne/spaceborne sensors are useful for large-scale biomass mapping, but their applications for sawgrass biomass quantification in the Everglades are limited. Smith et al. (2000) are the first and only researchers who attempted to assess sawgrass biomass in the Everglades using digital photography. In their work, profile-oriented images were acquired and transformed so that different colors could represent live, dead, and the absence of plant materials. Pixel counts were used to estimate biomass within a frame. This profile-oriented digital photography approach cannot be used for large-scale quantification of biomass and this imaging technique is also obsolete. To monitor the spatial and temporal variation of coastal mash biomass in the Everglades, spaceborne sensors are more ideal than airborne sensors.

1.3. Modeling methods for biomass estimation of coastal marshes

Remote sensing modeling is a crucial stage for biomass estimation, and the selection of estimation methods may largely impact the final results (Lu et al., 2014). Previous studies in biomass estimation of non-woody coastal marshes most commonly applied the traditional regression approaches such as partial least squares regression and multiple linear regression algorithms. These parametric regression algorithms assume that the relationship between the dependent (biomass) and independent (derived from remote sensing data) variables has an explicit model structure that can be specified by parameters. If the biomass is nonlinearly related to the remote sensing variables, these algorithms are likely to fail. Contemporary machine learning regression techniques have proven valuable to model the nonlinear relationship. They have been extensively used in forest biomass modeling, as reviewed by Lu et al. (2014), but applications in biomass estimation of non-woody coastal marshes are limited. There is a need to expand machine leaning modeling techniques into biomass mapping of coastal marshes as an alternative to parametric algorithms.

In this study, four nonparametric machine learning regression methods and one parametric algorithm were evaluated for sawgrass biomass modeling, including Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbor (k-NN), Artificial Neural Network (ANN), and Multiple Linear Regression (MLR). These algorithms may generate comparable results but different predictions, thus a combination of the predictions from each model will produce a more robust mapping result through an ensemble analysis. To the best of our knowledge, an ensemble analysis of the predictions from multiple models has never been conducted in biomass modeling, even though it has been widely used in image classification, as reviewed by Du et al. (2012). An ensemble analysis of the outputs from machine learning algorithms has achieved encouraging results for vegetation classification in the Everglades (Zhang, 2014, 2015; Zhang et al., 2016). Ensemble analysis was evaluated for biomass modeling in this study.

1.4. Biomass mapping methods

Research in the optical remote sensing of biomass has concentrated on pixel analysis by matching in-situ measurements with individual pixels to develop the biomass model, and generating the biomass map at the pixel level. In practice, coastal planners and researchers are less interested in a single pixel than a composition or a configuration of many pixels that comprise a meaningful landscape within a scene. Biomass of a patch/region is more useful. Object-based image analysis (OBIA) offers a promising approach for such applications. OBIA has been well developed and widely used in image classification, as reviewed by Blaschke (2010). However, its applications in plant biomass modeling and mapping are sparse (Addink et al., 2007). OBIA analyzes objects instead of pixels, which can reduce the "salt-and-pepper" effect in mapping heterogeneous landscapes and enhance the analysis accuracy in wetland environments (Dronova, 2015). Another benefit of OBIA is that it allows for segmenting one image into objects/segments at multiple scales, providing a hierarchical set of scaled representations (Zhang, 2016). In contrast, the pixels are uni-scale and represent a fixed area on the ground with a non-hierarchical and single value. It is expected that OBIA is a better solution to map sawgrass biomass than the pixel-based method.

1.5. Objectives of this study

Few efforts have been made to evaluate the capability of modern remote sensing techniques for mapping sawgrass biomass in the Everglades. A sawgrass biomass map in the coastal Everglades has never been generated and published. OBIA, machine learning, and ensemble analysis techniques have been extensively applied in image classification for various applications, but their applications in biomass modeling are limited. A combination of these techniques for biomass modeling and mapping is even more scarce. To this end, the primary objective of this study is to develop an object-based ensemble approach to quantify sawgrass aboveground biomass at multiple scales using Landsat data by integrating OBIA, machine learning, and ensemble techniques. Studies have shown that Landsat data can be used to derive cost-effective, accurate estimates of biomass for coastal non-woody marshes, but this has not been explored in the Everglades. The specific objectives of this study include: 1) to evaluate whether Landsat data is effective for sawgrass biomass monitoring in the coastal Everglades; 2) to examine whether machine learning modeling algorithms are better than the parametric algorithms; 3) to assess the benefits of object-based techniques for biomass modeling and mapping compared with the pixel-based method; and 4) to explore the potential of ensemble analysis in biomass quantification using remote sensing.

2. Study area and data

2.1. Study site

The study site is located in the coastal Everglades near the Turkey Point Nuclear Generating Station of Florida Power & Light (FPL) (Fig. 1). The site is the region where fresh water from the Everglades via the 50-feet wide and 20-mile long C-111 canal meets the salt water in Florida Bay. This area is an important transitional zone incubating a number of economically valuable crustaceans. The C-111 canal has modified the coastal wetlands by misdirecting fresh water flow since 1960s. Restoration of the natural flow of water to Florida Bay is the key goal of CERP. The C-111 restoration project in CERP seeks to fill the southern portion of the waterway and replace it with an east-west