

Mapping forest aboveground biomass with a simulated ICESat-2 vegetation canopy product and Landsat data

Lana L. Narine^{1§}, Sorin Popescu¹, Tan Zhou¹, Shruthi Srinivasan², Kaitlin Harbeck³

Narine L.L., Popescu S., Zhou T., Srinivasan S., Harbeck K., 2019. Mapping forest aboveground biomass with a simulated ICESat-2 vegetation canopy product and Landsat data. Ann. For. Res. 62(1): _-_.

Abstract. The assessment of forest aboveground biomass (AGB) can contribute to reducing uncertainties associated with the amount and distribution of terrestrial carbon. The Ice, Cloud and land Elevation Satellite-2 (ICESat-2) was launched on September 15th, 2018 and will provide data which will offer the possibility of assessing AGB and forest carbon at multiple spatial scales. The primary goal of this study was to develop an approach for utilizing data similar to ICESat-2's land-vegetation along track product (ATL08) to generate wall-to-wall AGB maps. Utilizing simulated daytime and nighttime ICESat-2 data from planned ICESat-2 tracks over vegetation conditions in south-east Texas, we investigated the integration of Landsat data and derived products for AGB model and map production. Linear regression models were first used to relate simulated photon-counting lidar (PCL) metrics for 100 m segments along ICESat-2 tracks to reference airborne lidar-estimated AGB over Sam Houston National Forest (SHNF) in south-east Texas. Random Forest (RF) was then used to create AGB maps from predicted AGB estimates and explanatory data consisting of spectral metrics derived from Landsat TM imagery and land cover and canopy cover data from the National Land Cover Database (NLCD). Using RF, AGB and AGB uncertainty maps produced at 30 m spatial resolution represented three data scenarios; (1) simulated ICESat-2 PCL vegetation product without the impact of noise (no noise scenario), (2) simulated ICESat-2 PCL vegetation product from data with noise levels associated with daytime operation of ICESat-2 (daytime scenario), and (3) simulated ICESat-2 PCL vegetation product from data with noise levels associated with nighttime operation of ICESat-2 (nighttime scenario). The RF models exhibited moderate accuracies (0.42 to 0.51) with RMSE values between 19 Mg/ha to 20 Mg/ha with a separate test set. The adoption of a combinatory approach of simulated ICESat-2 and Landsat data could be implemented at larger spatial scales and in doing so, ancillary data such as climatic and topographic variables may be examined for improving AGB predictions.

Keywords: Lidar, ICESat-2, AGB mapping, ATL08

Authors. ¹Texas A&M University, 534 John Kimbrough Blvd. WFES Building, Room 331, TAMU 2120, College Station, TX 77843-2120, USA | ²Texas

A&M Forest Service, 200 Technology Way, Suite 1281, College Station, TX 77845-3424, USA | ³NASA Goddard Space Flight Center, NASA Goddard Space Flight Center, Greenbelt, MD, United States 20771, USA.

⁵Corresponding author: Lana L. Narine (ll_narine@tamu.edu)

Manuscript received January 4, 2019; revised February 20, 2019; accepted February 27, 2019; online first March 6, 2019.

Introduction

Accurate and spatially complete assessments of AGB are indicative of the extent to which forests contribute to the global carbon budget and can reduce uncertainties with the quantity and distribution of terrestrial carbon stocks (Houghton 2007, Houghton et al. 2007, Goetz & Dubayah 2011, Montesano et al. 2015). AGB carbon is primarily estimated using field inventory data where direct measurements such as tree diameters, are converted to AGB using biomass expansion factors or allometric regression equations (Brown 2002). These biomass regression equations are developed using allometric data collected from a sample of trees that have been destructively harvested and also usually over limited areas (Brown 2002, Popescu 2007). Combined with field measurements of AGB, remote sensing data can be used to reliably estimate AGB over multiple spatial scales (Popescu 2007, Nelson et al. 2017). Spaceborne light detection and ranging (lidar) instruments, in particular, are capable of providing measurements on a global scale and across areas that are otherwise difficult to access or where cost is prohibitive (Nelson et al. 2017). Reliable estimations of canopy height, basal area, AGB and aboveground carbon have been derived with spaceborne lidar (Nelson et al. 2017). While spatially continuous, global measurements from lidar are not yet possible, optical remotely sensed data can be integrated to generate spatially complete coverage of AGB and forest carbon at such scales (Hu et al. 2016).

The Ice, Cloud, and land Elevation Satellite (ICESat) operated from 2003 to 2010 and carried the Geoscience Laser Altimeter System

(GLAS), a waveform lidar system, which provided global elevation data during the course of the mission (Zwally et al. 2002). Its primary objective was to capture elevation changes of the earth's polar ice sheets and reduce uncertainties with the ice sheet mass balance (Zwally et al. 2002). GLAS produced 1064 nm and 532 nm laser pulses at a frequency of 40 Hz and illuminated a spot on the ground measuring approximately 60 m in diameter every 172 m in the along-track direction (Zwally et al. 2002). The spaceborne data collected by GLAS offers multidisciplinary benefits, which includes global topography and vegetation canopy heights (Zwally et al. 2002, Schutz et al. 2005). The literature demonstrates the utility of GLAS data for estimating forest heights (Harding & Carabajal 2005, Lefsky et al. 2007, Simard et al. 2011, Baghdadi et al. 2014) and AGB (Lefsky et al. 2005; Nelson et al. 2017). Furthermore, to overcome the spatial discontinuity of the lidar measurements, several studies have demonstrated approaches that integrate data from spaceborne multispectral sensors to generate wall-to-wall maps of AGB (Duncanson et al. 2010, Hudak et al. 2002; Chi et al. 2015, Hu et al. 2016). Essentially, the use of satellite imagery, such as data from Landsat sensors and from the moderate-resolution imaging spectroradiometer (MODIS) has been widely demonstrated across large-scale mapping studies.

With the recent launch of ICESat-2 on September 15th, 2018 (NASA 2017), up-to-date vegetation data at near-global scales and new prospects for vegetation mapping, will be available. This follow-on 3-year mission has been designed to overcome some of the limitations associated with ICESat to provide

improved sampling and increased spatial coverage, as described in Markus et al. (2017). Specifically, ICESat-2's photon-counting lidar (PCL) instrument, the Advanced Laser Altimeter System (ATLAS), splits a 532 nm laser pulse into three pairs of beams spaced 3.3 km apart with a 90 m pair spacing (Gwenzi & Lefsky 2014, Markus et al. 2017). It operates at an increased repetition rate of 10 kHz, facilitating denser sampling and producing smaller footprints than GLAS, measuring 14 m and up to 17 m every 0.7 m along the ground track (Leigh et al. 2015, Markus et al. 2017). Since vegetation has lower reflectance than ice surfaces at the 532 nm wavelength, approximately 1/3 to 1/9 of the photon returns from ice and snow surfaces is expected for terrestrial surfaces (Neuenschwander & Magruder 2016). The number of signal photons per transmitted pulse will range between 0 and 3 over vegetated surfaces (Neuenschwander & Magruder 2016). However, differentiating signal photons from noise photons represents a challenge with photon counting systems like ATLAS, so effective filtering algorithms are crucial to deriving accurate estimates (Glenn et al. 2016). Noise levels also vary with operation times of ICESat-2, with less background noise associated with nighttime operation of ATLAS and increased potential for more accurate canopy height retrievals than from daytime retrievals (Popescu et al. 2018).

One of the ICESat-2's data products is the Land Water Vegetation Elevation or ATL08 product, which will provide terrain and canopy heights for non-polar regions (Neuenschwander et al. 2017). Terrain and height estimates will be provided at a fixed segment size of 100 m along the ground track. The ATL08 product will complement data from other space-based missions, like the Global Ecosystem Dynamics Investigation (GEDI) which was launched on December 5th, 2018 (NASA 2019) as well as data from optical sensors. The examination of approaches for AGB estimation and mapping could allow for development of an ap-

propriate methodological framework for use as soon the ATL08 product becomes available. However, studies pertaining to the use of ICESat-2 data, especially its vegetation product for characterizing vegetation, are limited. For instance, Montesano et al. (2015) simulated ICESat-2 data and indicated difficulty in capturing vegetation structure where vegetation is sparse. An analysis of data from the Multiple Altimeter Beam Experimental Lidar (MABEL) instrument, ICESat-2's demonstrator instrument (Glenn et al. 2016), highlighted a synergistic approach between ICESat-2 and Landsat for improving AGB estimates. Gwenzi et al. (2016) indicated similar performance between ICESat and ICESat-2 using MABEL data, in estimating heights for savanna vegetation.

The aim of this study was to examine the use of ICESat-2 data, for generating wall-to-wall AGB coverage. The conceptual approach to AGB mapping has been highlighted in previous studies with GLAS data (Duncanson et al. 2010; Chi et al. 2015). However, considerable differences between the two instruments, such as measurement concept, spatial coverage and data products (Markus et al. 2017) warrant insights into approaches for AGB mapping with expected data from ICESat-2. In this study, a synergistic approach with data similar to what will be provided by ICESat-2 and Landsat products was explored. With expected differences in noise levels associated with daytime and nighttime operation of ATLAS and associated impacts on canopy height estimation (Popescu et al. 2018), an examination of AGB mapping under different data scenarios are also included in this study. The primary goal of this study was to create wall-to-wall AGB maps at 30-m spatial resolution using simulated PCL-estimated AGB from known ICESat-2 track locations over Sam Houston National Forest (SHNF) in Texas and predictor variables from Landsat data. Three AGB maps were created from the following data scenarios: (i) Simulated ICESat-2 PCL vegetation product

without the impact of noise (no noise scenario), (ii) Simulated ICESat-2 PCL vegetation product from data with noise levels associated with daytime operation of ICESat-2 (daytime scenario), and (iii) Simulated ICESat-2 PCL vegetation product from data with noise levels associated with nighttime operation of ICESat-2 (nighttime scenario). The same modeling technique was applied to generate AGB maps for the three scenarios with spatially explicit maps of model uncertainty produced for each corresponding AGB density map. ICESat-2 will provide global-scale sampling of forest resources and could potentially be leveraged to provide rich insight about the earth's forests. At the end of its mission, ICESat-2 data will be used to generate ATL18 products which consist of gridded terrain and canopy maps (Neuenschwander et al. 2017). This study ultimately serves to support upcoming AGB mapping with ICESat-2 and advance our understanding of the data for vegetation studies.

Materials and methods

Study area

The study area is located in SHNF in southeast Texas, USA (Latitude 30° 42' N, Longitude 95° 23' W) within the Pineywoods ecoregion and consists of approximately 48 km² of land. Summer temperatures average 28°C and the average winter temperature is around 12°C (USDA Forest Service 2018). The site consists of gentle slopes and elevations from 62 m and 105 m with a mean elevation of 85 m (Popescu 2007). Approximately 80% of the forests are classified as evergreen forest (Homer et al. 2015; MRLC 2017), which include Loblolly pine (*Pinus taeda*) plantations and old growth Loblolly pine stands (Popescu 2007), while common hardwoods found on floodplains and drier uplands include White Oak (*Quercus alba*) and Southern Red Oak (*Quercus falca*). Planned ICESat-2 track locations for the study site, provided by the mission's Science Defi-

nition Team (SDT) (Neuenschwander et al. 2017), are shown in Figure 1.

Simulated ICESat-2 data scenarios

Airborne lidar data collected in November 2010 were used to simulate ICESat-2 data along known ICESat-2 tracks over SHNF. The data were acquired along 12 flight lines in the north-south direction and 19 lines flown from east to west, at around 600 m above-ground-level (AGL) and has a point density of 4 points per m². The airborne lidar data and simulated ICESat-2 footprints with a center-to-center spacing of 70 cm along each ground track were used as input for simulation of the data scenarios. The simulation algorithm, described in detail in Neuenschwander & Magruder (2016), was developed by the ICESat-2 SDT. To summarize, four main steps were implemented (Neuenschwander & Magruder 2016): (1) assemble heights from discrete return points within a 7 m radius of footprint centers to generate a pseudo-waveform (Blair & Hofton 1999), (2) construct a height vector for each footprint, (3) randomly determine the number of photons (x) to sample per shot based on design cases developed by the mission's instrument team, and (4) randomly sample the vector of heights x times, weighted by the pseudo-waveform. The design cases developed by the mission's instrument team indicate the expected number of photons per footprint vary by vegetation type and can range from 0 to 3 returns per pulse (Neuenschwander & Magruder 2016). The mean number of signal photons per footprint for SHNF was modeled at 1.9, based on the ATLAS performance model for temperate forests (Martino 2010). To complete the simulation, background noise, representing anticipated solar background noise or from the atmosphere was added to the data. Since solar background noise is expected to be pronounced during daytime operation of ICESat-2 (Degnan 2002), the datasets were generated to reflect the different expected noise levels: (1) daytime scenario with expected noise

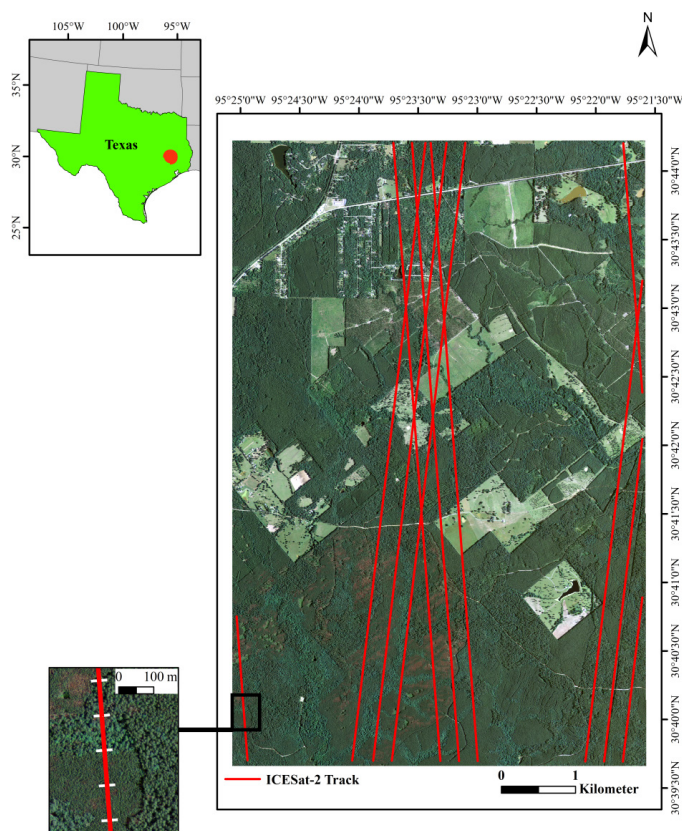


Figure 1 ICESat-2 tract locations overlaid on 2010 National Agriculture Imagery Program (NAIP) aerial imagery for the study area within SHNF, Texas (inset map, upper left corner) with demarcation of 100 m segments along-track on inset map, lower left corner.

levels for daytime operation of ICESat-2, and (2) nighttime scenario with noise levels based on night operation of ICESat-2. Noise photons are not discernable from signal photons so the implementation of effective processing algorithms is critical in order to derive accurate forest measurements from PCL data. Noise filtering and photon classification algorithms (Popescu et al. 2018) were applied to the daytime and nighttime scenarios to derive top of canopy and ground estimates. As presented in Popescu et al. (2018), noise filtering entailed a multi-level approach for minimizing noise photons and moving overlapping windows

and cubic spline interpolation were used to classify the remaining photons. In terms of accuracy of the algorithms, Popescu et al. (2018) reported average RMSE values of 1.83 m and 2.80 m for estimated ground elevation and 2.70 m and 3.59 m for canopy height estimations with the nighttime and daytime scenarios respectively. Following the application of the noise filtering and photon classification algorithms for daytime and nighttime scenarios investigated in this study, estimated canopy heights were then retrieved by subtracting estimated ground from top of canopy values.

The airborne lidar data for the study area was also processed to obtain above-ground-level heights and used for PCL simulation without the addition of background noise (no noise scenario). In total, three simulated PCL datasets (data scenarios) were generated for each planned ICESat-2

track over SHNF.

Reference AGB

Reference AGB estimates were calculated from a canopy height model (CHM) derived from airborne lidar data for SHNF. The CHM was used as input in a lidar software application called TreeVaW, to extract forest inventory parameters for individual trees (Popescu et al. 2003, Popescu & Wynne 2004). TreeVaW uses a variable window technique with local maximum focal filtering to extract

tree heights, where the window size is based on the relationships between tree heights and crown widths from field inventory data (Popescu 2007). Popescu & Wynne (2004) indicated that lidar measurements processed with this technique explained 97% of the variance associated with mean height of dominant pine trees. For SHNF, ground measured tree heights and corresponding crown diameters from a total of 705 pine trees and 603 deciduous trees from field inventories carried out in 2004 and 2009 were used to obtain TreeVaW coefficients to subsequently derive tree locations and measurements. The crown diameter of an individually located tree in the CHM was also estimated in TreeVaW, which involves averaging two values measured along perpendicular profiles from the center of a tree top by fitting a fourth-degree polynomial on each profile (Popescu et al. 2003, p. 564). Details of the processing approach implemented in TreeVaW are presented in Popescu et al. (2003) and Popescu & Wynne (2004). Output consisting of estimated diameter at breast height (dbh) measurements, were used to estimate tree-level AGB using a generalized biomass regression equation (eq. 1):

$$AGB_{(kg)} = \exp(\beta_0 + \beta_1 \ln(dbh)) \quad (1)$$

where dbh - the diameter at breast height (cm) and β_0 and β_1 - the parameters for the species group (Jenkins et al. 2003, p. 20). AGB for pines were calculated using the parameters for the “pine” species groups, with $\beta_0 = -2.5356$ and $\beta_1 = 2.4349$. To calculate AGB for deciduous trees, parameters for the “hard maple/oak/hickory/beech” group were used, with $\beta_0 = -2.0127$ and $\beta_1 = 2.4342$ (Jenkins et al. 2003, p. 20).

AGB estimation from simulated PCL-derived metrics

For each data scenario, 100 m segments along ICESat-2 profiles were extracted and maxi-

mum height, mean height, height percentiles, canopy cover and canopy density were calculated for each segment. A segment measuring 100 m in the along-track direction was chosen in order to be consistent with the format of the planned ATL08 product from ICESat-2 (Neuenschwander et al. 2017). A total of 121 segments over the study site were used, where one-third was randomly assigned for model testing and the remaining 85 segments were used for developing the models for estimating AGB. Linear regression models were used to relate the simulated PCL metrics for a subset of the segments to spatially coincident, airborne-lidar estimated AGB and the resulting models were evaluated with a separate test set. Details of the methodology used to estimate AGB and canopy cover from simulated PCL metrics are provided in Narine et al. (2019).

Using resulting AGB models, AGB density (Mg/ha) was estimated for the segments over SHNF and then applied to 30 m pixels to match Landsat TM pixels. Since different segment lengths traverse pixels, AGB was calculated based on the portion of a 100 m segment across a pixel and the value was extrapolated to represent the pixel size using the estimated area of a segment across a pixel. Pixels with segments measuring less than 7 m in a pixel were excluded from analysis and average pixel AGB was calculated in instances where there were two segments (parts) of equal lengths. The steps for assigning AGB to pixels were repeated for each scenario using the corresponding prediction equation developed from linear regression analysis.

Mapped predictors

Landsat data is freely available and offers global coverage (Avitabile et al. 2012). Landsat data, specifically, Landsat 5 Thematic Mapper (TM) was chosen for use in this study to avoid temporal mismatch with the simulated data and reference AGB from airborne lidar data acquired in November, 2010. The data is

provided at spatial resolution of 30 m and has 7 spectral bands, which include blue, green, red and near infra-red (NIR). A Landsat cloud-free scene encompassing the study site, from path/row 26/39, acquired in November, 2010 was downloaded from U.S. Geological Survey (USGS) Earth Explorer in GeoTIFF format and re-projected to UTM, WGS84, Zone 15N. The image was processed in ENVI to top-of-atmosphere (TOA) reflectance and vegetation indices were computed at the 30 m pixel size (Table 1). Land cover and canopy cover maps from the 2011 National Land Cover Database (NLCD) (Homer et al. 2015) were also downloaded and clipped to the extent of the study site for use as predictor variables. Land cover and canopy cover maps were georeferenced in the same projection as the Landsat TM image.

AGB mapping

The regression tree method, Random Forest (RF) (Breiman 2001), was used to model AGB for the three data scenarios using spectral metrics from Landsat 5 TM, landcover and canopy cover and estimate AGB for areas not overlaid by ICESat-2 tracks. RF is a nonparametric modeling technique and an established approach for mapping AGB using multisource data (Baccini et al. 2004, Houghton et al. 2007, Chi et al. 2015). Using a bootstrap sample of the training data, the best split is made at the root node using a random sample of the predictors (Liaw & Wiener 2002). Another sample of the variables is taken at the other node

and the process is repeated until the regression tree grows as large as possible. The process is repeated with a new bootstrap sample and the data is predicted by averaging the predictions of all trees (Liaw & Wiener 2002). The un-sampled training data at each bootstrap iteration, called out-of-bag or OOB data, is used to generate OOB model predictions which are aggregated to calculate the error rate (Liaw & Wiener 2002). RF was carried out with the ModelMap R package (Freeman et al. 2018) which calls randomForest R package (Freeman et al. 2018). ModelMap was used to validate the model with OOB predictions on the training data and also with a separate test set and apply the model to create the final AGB map for each scenario (Freeman et al. 2018). About 70% of the data was randomly assigned to the training dataset and remaining 30% was allotted to the test set and used for model evaluation. As a result, the training and test set for each scenario consisted of 1448 and 620 30-m pixels respectively, equating to approximately 3% and 1% of the pixels which constitute the study area. A total of 1000 regression trees were built with the training data for each scenario and results were used to assign AGB values to each 30 m pixel. The final AGB maps were masked using the land cover map, where AGB density in non-forest areas (NLCD classes: water, developed, barren, grassland/herbaceous, pasture and emergent herbaceous wetlands) was set to 0 Mg/ha.

Since trees in a RF model are independent, the standard deviation of individual-tree pre-

Table 1 Mapped predictor variables for RF regression

Variable	Description
Spectral Metrics from Landsat 5 TM	
Normalized Difference Vegetation Index (NDVI)	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$
Enhanced Vegetation Index (EVI)	$2.5 \cdot ((\text{NIR} - \text{Red}) / (\text{NIR} + 6 \cdot \text{Red} - 7.5 \cdot \text{Blue} + 1))$
Soil Adjusted Vegetation Index (SAVI)	$((\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + 0.5)) \cdot (1.5)$
Modified Soil Adjusted Vegetation Index (MSAVI)	$(2 \cdot \text{NIR} + 1 - \text{sqrt}((2 \cdot \text{NIR} + 1)^2 - 8 \cdot (\text{NIR} - \text{Red}))) / 2$
Land Cover	National Land Cover Database 2011 (NLCD 2011)
Canopy Cover	NLCD 2011 US Forest Service Tree Canopy Cover

dictions can be calculated at the pixel-level to produce a measure of prediction variability (Freeman et al. 2016). According to Freeman et al. (2016), standard deviation of predictions for trees can be used as a measure of uncertainty in resulting maps. High uncertainty values indicate a lack of agreement among trees (Freeman et al. 2016, p. 15), with some trees predicting high AGB and others low AGB values while low uncertainty values translate to agreement in predictions from individual trees for a pixel. RF uncertainty maps corresponding to final AGB maps were generated.

Results were also compared with AGB estimated from airborne lidar data over the study. The reference AGB estimates derived from airborne lidar data were aggregated by summation at a 30 m resolution and pixel AGB values were converted to AGB density (Mg/ha). Cells spatially coincident with the original test datasets were then selected and predicted AGB were compared with the reference AGB pixels. AGB predictions for the three data scenarios were compared using RMSE and R^2 metrics.

Results

Estimated AGB from simulated PCL metrics

Details of results from regression analysis used to relate simulated PCL height metrics, canopy cover and canopy density for 100 m segments to reference airborne lidar-derived AGB are presented in Narine et al. (2019). To summarize, models had high accuracies and yielded RMSE values of 19.16 Mg/ha, 25.35 Mg/ha, and 19.23 Mg/ha for the no noise scenario, daytime scenario and nighttime scenario respectively. The prediction equation for the simulated dataset without noise used mean height and a canopy density variable for 15 m to 20 m height bin and explained 79% of the variance in airborne lidar-estimated AGB. The absence of noise photons and the applied photon detection rate for the forest being studied rendered this dataset the best case scenar-

io. Given that background noise represents a challenge for photon counting systems (Glenn et al. 2016) estimation models associated with daytime and nighttime operation times offer more insight about the potential AGB estimation from ICESat-2 data. AGB models with the daytime and nighttime scenarios used the same variables; the 10th and 90th height percentiles and canopy cover calculated as the proportion of returns above 4.6 m. With the daytime scenario, the variables explained 63% of the variance in airborne lidar-derived AGB while the same variables calculated with the simulated nighttime dataset yielded a R^2 of 0.79.

RF predicted AGB

Using the simulated PCL-estimated AGB as the dependent variable and spectral metrics, land cover and canopy cover, the RF regression tree models explained 49%, 39% and 47% of the variance with OOB data for the no noise, daytime and nighttime scenarios respectively. RMSEs estimated using OOB testing were 20.59 Mg/ha, 19.98 Mg/ha and 19.25 Mg/ha. With the test data, the predictive abilities of models, represented by the R^2 values, were 0.51, 0.42 and 0.49. Overall, the models tended to underestimate cells with high biomass (Figure 2). With test data, the RMSE values for the no noise and daytime scenarios were 19.69 Mg/ha and 19.67 Mg/ha, and slightly improved for the nighttime scenario, with a RMSE of 19.31 Mg/ha. The relationship between RF predicted AGB and AGB estimated from airborne lidar data over the study site yielded lower R^2 values and higher RMSEs and similar results for the three scenarios. When rounded to the nearest whole number, the RMSE was 35 Mg/ha with the no noise, daytime and nighttime scenarios while the variance explained was 46% with the no noise and nighttime scenarios and 45% with the daytime scenario (Table 2). The most important predictor of AGB for the three data scenarios was canopy cover followed by NDVI and land cover. Variable importance was almost split between NDVI and land cover

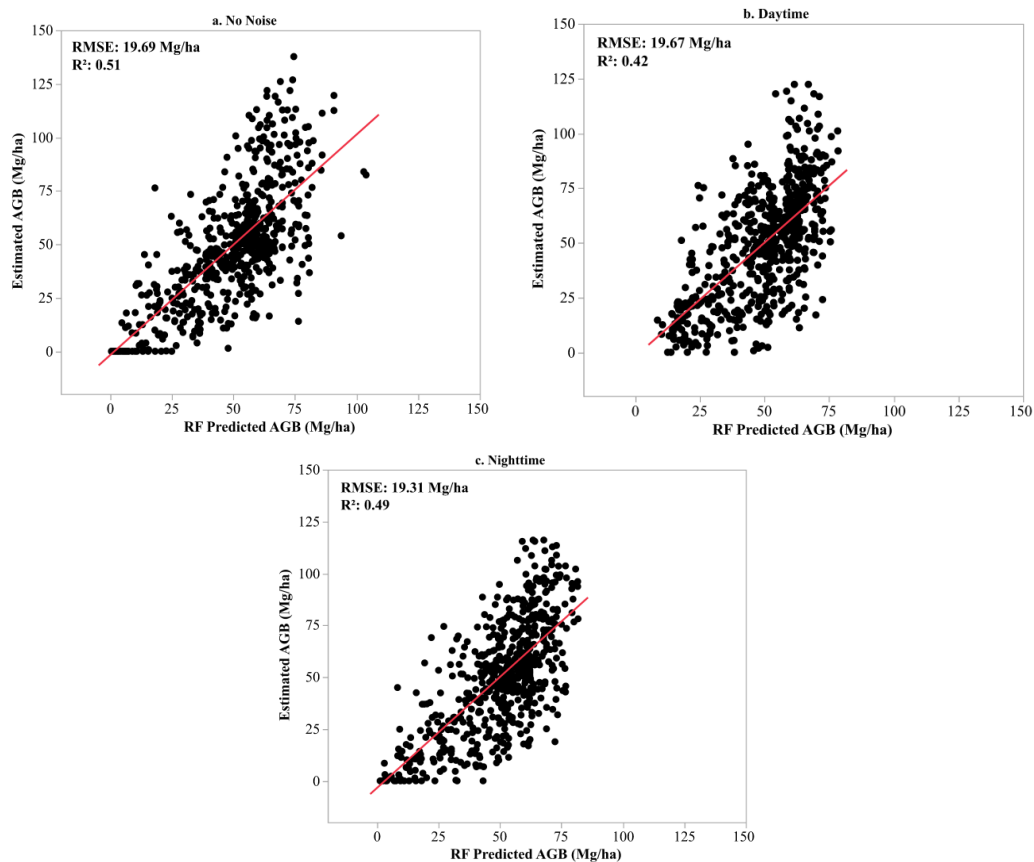


Figure 2 (a) Simulated PCL AGB estimated from linear regression vs RF predicted AGB with test data for the no noise scenario; (b) Simulated PCL AGB estimated from linear regression vs RF predicted AGB with test data for the daytime scenario; (c) Simulated PCL AGB estimated from linear regression vs RF predicted AGB with test data for the night time scenario.

while SAVI and MSAVI were the least important predictors of AGB.

Average AGB values from final AGB maps (Figures 3-5) generated with the RF regression models were 53.05 Mg/ha, 50.73 Mg/ha and 52.39 Mg/ha for the no noise, daytime and nighttime scenarios respectively. In comparison, average AGB of forests in the training data were 52.09 Mg/ha, 51.21 Mg/ha, and 52.24 Mg/ha. With this data, AGB ranged from 0 Mg/ha to 150.64 Mg/ha for the no noise scenario, 0 Mg/ha to 139.61 Mg/ha, and 0 Mg/ha to 119.75 Mg/ha for the daytime and nighttime scenarios respectively. AGB predictions for 30

m pixels were within the range of the data but tended to underestimate AGB. AGB density predictions of forests in the study site ranged from 0.72 Mg/ha to 105.41 Mg/ha for the no noise scenario, 6.62 Mg/ha to 83 Mg/ha and 1.97 Mg/ha to 83.55 Mg/ha for the daytime and nighttime scenarios respectively.

Visual assessments of AGB maps revealed similarities in the spatial distribution pattern of predicted AGB for the three scenarios (Figures 3-5). However, the daytime and nighttime scenarios appeared to have more areas with AGB densities ranging from 20-40 Mg/ha than the no noise scenario while the map

Table 2 Test set error statistics from RF models predicting AGB and relationships with aggregated airborne lidar-derived estimates of AGB under three scenarios; no noise, daytime and nighttime scenarios.

Scenario	RF Model Performance with Test Set		Relationship with Reference AGB	
	R^2	RMSE (Mg/ha)	R^2	RMSE (Mg/ha)
No Noise	0.51	19.69	0.46	34.99
Daytime	0.42	19.67	0.45	35.27
Nighttime	0.49	19.31	0.46	34.93

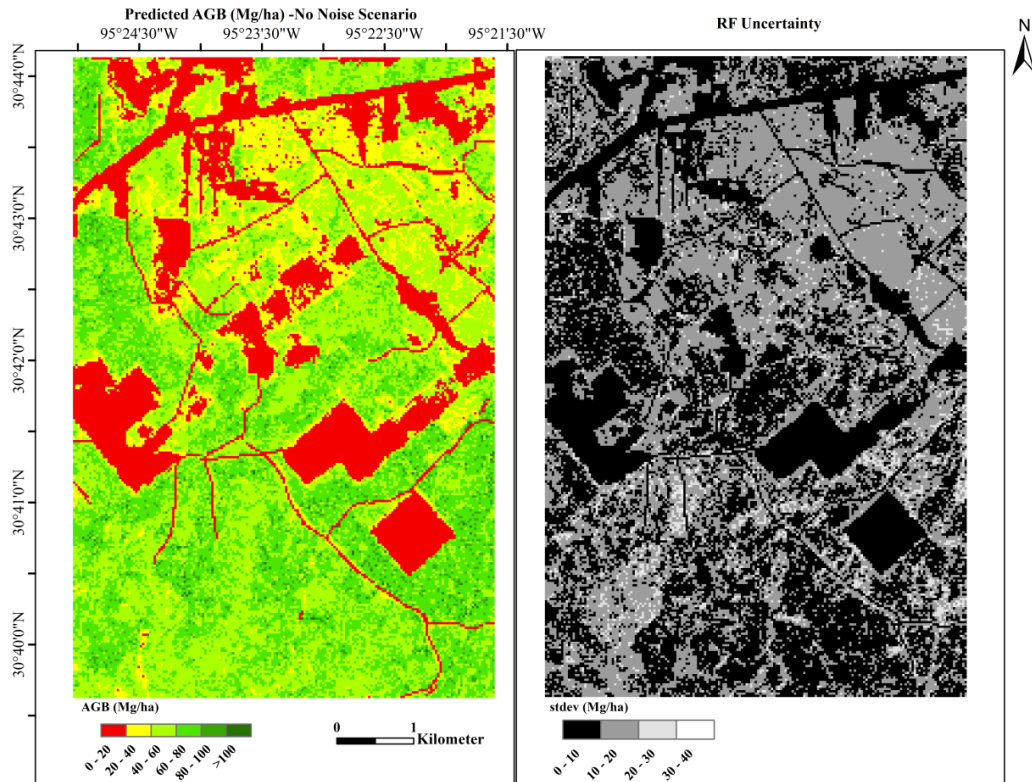


Figure 3 (a) RF predictions of AGB density at a 30 m spatial resolution for the no noise scenario; (b) RF uncertainty at a 30 m spatial resolution for the no noise scenario.

for the daytime scenario exhibited the greatest predominance of relatively low AGB values. To demonstrate, predicted AGB estimates with the three scenarios for a highlighted portion of the study area, is shown on Figure 5. Within the highlighted extent, predicted AGB for the daytime and nighttime scenarios exhibited lower AGB estimates than the no noise scenario. Compared to the other scenarios, lower AGB

estimates were most prevalent with the daytime scenario, represented by the 20-40 Mg/ha range. Overall, AGB maps correspond to vegetation trends present with higher AGB in the southern portion of the study site primarily occupied by mature pines and lower biomass occupied by young pine stands. At the pixel level, predictions of AGB from individual trees that var-

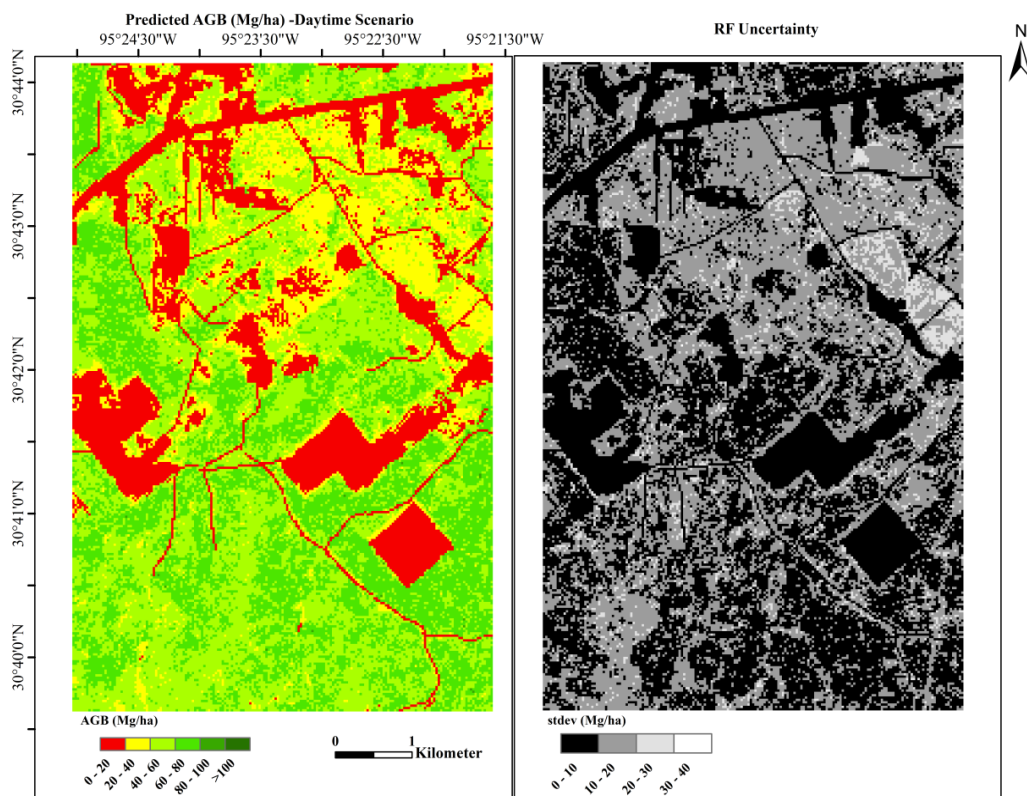


Figure 4 (a) RF predictions of AGB density at a 30 m spatial resolution for the daytime scenario; (b) RF uncertainty at a 30 m spatial resolution for the daytime scenario.

ied greatly led to a high standard deviation of the predictions. Corresponding RF uncertainty maps (Figures 3-5) for the scenarios highlight pixels with particularly large differences in individual tree predictions, especially for the daytime scenario. RF uncertainties highlighted in Figure 5 emphasize the predominance of high standard deviation of AGB predictions for the daytime scenario, compared to the other data scenarios. In this area, the 10-20 Mg/ha range was most prevalent with the no noise scenario followed by the nighttime scenario. In comparison, there were substantially more pixels within the 20-30 Mg/ha range, indicative of greater uncertainty, for the daytime scenario (Figure 5).

Discussion

The mapping of AGB has been identified as an approach to advance the knowledge about the terrestrial aspect of the carbon cycle (Le Toan et al. 2011). Consequently, it is necessary to develop approaches for characterizing the spatial distribution of AGB and forest carbon. ICESat-2 presents an opportunity to obtain large-scale coverage about vegetation through the collection of data along transects on the earth's surface. Despite the lack of spatially complete data that will be measured from this source, complete coverage offered by Landsat or MODIS sensors may be leveraged to achieve comprehensive estimates of forest attributes. This study presents one approach for

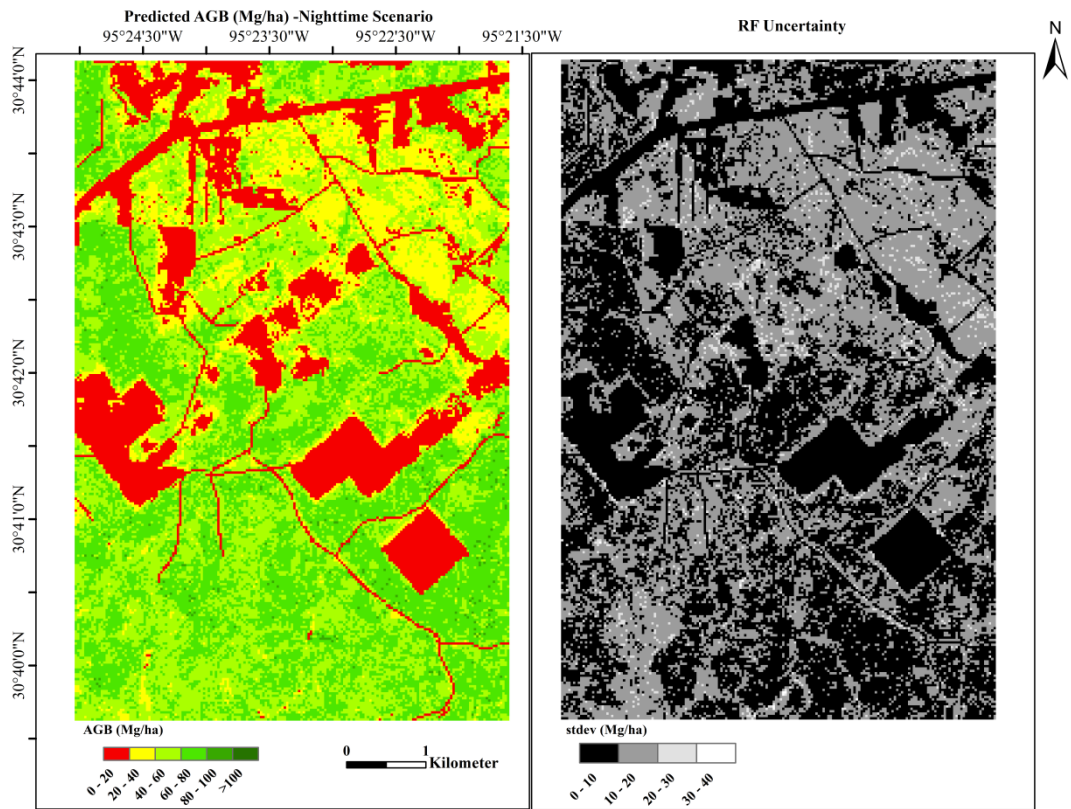


Figure 5 (a) RF predictions of AGB density at a 30 m spatial resolution for the nighttime scenario; (b) RF uncertainty at a 30 m spatial resolution for the nighttime scenario.

deriving wall-to-wall coverage of AGB from ICESat-2. Two sets of models were used; one for modeling the relationship between simulated PCL metrics for 100 m segments and reference airborne lidar-derived AGB and another set for relating mapped predictor variables to simulated PCL-estimated AGB. Using nonparametric regression models, AGB density was extrapolated from 3% of the 30 m pixels which comprise the study site to achieve spatially explicit AGB values. While the study area was focused on the SHNF, this approach is applicable to scaling up to larger spatial extents. The use of 30 m map resolution is particularly useful for upscaling purposes with the availability of spatially comprehensive data, such as Land-

sat imagery and NLCD products. In doing so, other variables may also be investigated for improving AGB estimates. For example, information from optical imagery such as Tasseled cap indices, and ancillary data such as elevation and slope may be incorporated for regional scale mapping (Zald et al. 2016). Additionally, vegetation indices from multi-date satellite imagery could be investigated when upscaling to larger extents. To demonstrate, Li et al. (2015) highlighted the use of vegetation indices from multi-date imagery during the growing season for achieving more reliable results for regional AGB estimation. It is also important to note that coarser map resolutions may provide good accuracies. To exemplify, Deo et al. (2018) in-

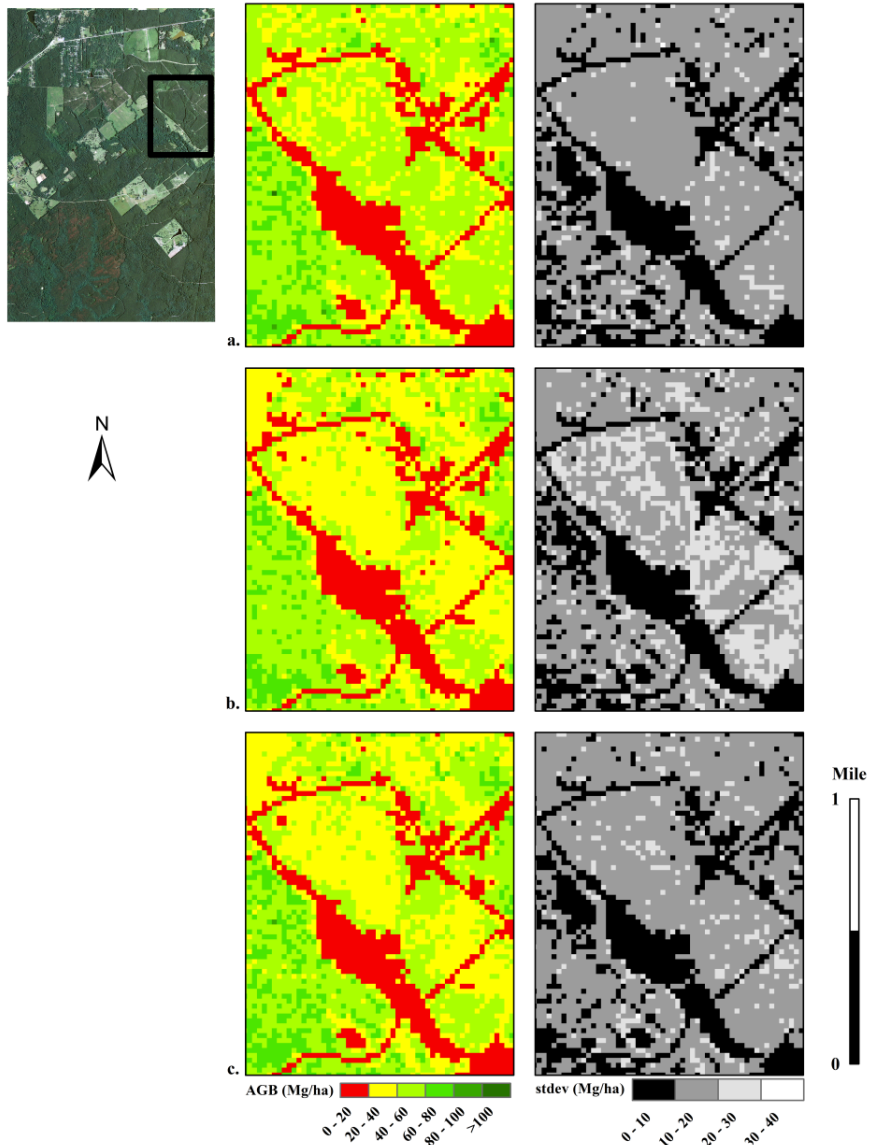


Figure 6 (a) RF predictions of AGB density at a 30 m spatial resolution for the nighttime scenario; (b) RF uncertainty at a 30 m spatial resolution for the nighttime scenario.

investigated the impact of spatial resolution on the accuracy of regional scale AGB maps encompassing the eastern USA from Landsat-derived parameters and concluded that grid sizes up to 1 km provided reasonable accuracies.

In this study, validation statistics for the AGB models indicated RMSEs ranging from 69.34 Mg/ha for the 30 m grid size, up to 80.78 Mg/ha with the 1000 m grid size but with similar correlation (r) results between predictions and

the reference data (Deo et al. 2018).

Canopy cover has proven to be a valuable predictor of AGB in this study. Canopy cover calculated from simulated PCL segments was a significant predictor of airborne lidar-derived AGB and results from the RF models produced in this study indicated NLCD canopy cover as the most important predictor of AGB, regardless of the data scenario. Alternative approaches to generating AGB maps may include extrapolating canopy cover estimated in the ATL08 product and then using this as a predictor variable in RF models. A similar approach with GLAS parameters is presented in Hu et al. (2016), where RF was used to generate spatially continuous GLAS parameters and then combined with mapped predictors including NDVI and landcover, to estimate global AGB at a 1-km resolution. In the case of extrapolating canopy cover and then predicting AGB, the possibility for simultaneously producing maps of these parameters will be beneficial as both represent fundamental attributes of forest vegetation structure.

An advantage of using RF for modeling is the possibility of using OOB as part of model evaluation (Freeman et al. 2016). An evaluation of RF model performance for the no noise, daytime and nighttime scenarios with a separate test set yielded similar results to those provided with OOB data; R^2 values were 0.51, 0.42, and 0.49, compared to 0.49, 0.39, and 0.47 with OOB testing. RMSE values also ranged between 19 Mg/ha and 21 Mg/ha with the OOB and independent test set. Overall, results from both sources emphasized the ability of the nighttime scenario to out-perform the daytime scenario. RF also provided an indicator of uncertainty in the resulting maps, estimated as the variability of predictions from 1000 independent trees used to estimate pixel AGB. A comparison of the AGB and AGB uncertainty maps for the three scenarios highlighted areas with lower AGB predictions and corresponding higher uncertainty ranges for the daytime versus the nighttime and no noise scenarios. Factors that may contribute to final

AGB uncertainty include allometric models for reference AGB estimation, input data errors (e.g. NLCD, Landsat reflectance) and methods used for upscaling AGB to the pixel size for modeling. Prospective studies could involve a comprehensive uncertainty analysis of AGB estimates predicted from actual ICESat-2 data to ascertain error originating from different sources.

GEDI commenced its two-year mission in November, 2018 and will provide data for characterizing three-dimensional vegetation structure at a global scale (NASA 2019). GEDI is designed specifically to provide data for characterizing three-dimensional vegetation structure (NASA 2019). While ICESat-2 will collect measurements between 88° north and south latitudes during its three-year duration (Markus et al. 2017), GEDI will provide waveform observations between 52° north and 52° south latitudes (Marselis et al. 2016). Hence, the adoption of a synergistic approach between ICESat-2 and GEDI may support a more accurate quantification of forest attributes, and specifically forest AGB. Future research should investigate potential synergies between ICESat-2 and GEDI for producing an AGB product.

Conclusions

Up-to-date and spatially explicit assessments of AGB density can support the monitoring of forest carbon, contributing to reduced uncertainties with the carbon budget and an improved understanding of changes in terrestrial carbon storage. In this study, we present an approach for mapping AGB with ICESat-2 which can be summarized as follows: (i) simulating and processing of ICESat-2 data to generate three data scenarios in a format similar to the ATL08 product, (ii) developing relationships between the simulated vegetation product data and airborne-lidar derived AGB, and then between predicted AGB and mapped predictors consisting of spectral metrics from Landsat 5 TM,

landcover and canopy cover using RF, and (iii) mapping AGB at 30-m spatial resolution and producing a corresponding measure of uncertainty using the resulting RF models. Findings highlight canopy cover as the most important predictor of AGB and indicate similarities in the predictive capabilities of RF models for the three settings analyzed, and especially for the nighttime scenario which outperformed the daytime scenario. Overall, the methodology is conducive to achieving wall-to-wall AGB coverage at much larger extents and in doing so, other variables such as topographic, climatic, and spectral indices may be considered for improving AGB estimates. Efforts focused on the contributions of error from different sources would also be needed to ascertain final AGB uncertainties. In addition, the recent launch of GEDI will advance the capability of assessing AGB, thus highlighting the need for investigations that support a synergistic approach between ICESat-2 and GEDI to characterize AGB and other forest attributes. Furthermore, while spatial coverage from both ICESat-2 and GEDI are not possible, the integration of spatially comprehensive data, like Landsat optical imagery, can potentially facilitate a more comprehensive and accurate AGB product.

Acknowledgments

This work was supported by NASA under Grant Number NNX15AD02G.

References

- Avitabile V., Baccini A., Friedl M.A., Schmillius C., 2012. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sensing of Environment* 117: 366-380. DOI:10.1016/j.rse.2011.10.012
- Baccini A., Friedl M.A., Woodcock C.E., Warbington R., 2004. Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters* 31: 1-4. DOI: 10.1029/2004gl019782
- Baghdadi N., le Maire G., Fayad I., Bailly J.S., Nouvellon Y., Lemos C.; Hakamada, R., 2014. Estimation of forest height and aboveground biomass from ICESat-2/GLAS data in eucalyptus plantations in Brazil. In 2014 IEEE International Geoscience and Remote Sensing Symposium, 13-18th July 2014, Quebec City, QC, Canada. IEEE, pp. 725-728. DOI: 10.1109/igarss.2014.6946526
- Blair J.B., Hofton M.A., 1999. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters* 26: 2509-2512. DOI: 10.1029/1999gl010484
- Breiman L., 2001. Random forests. *Machine Learning* 45: 5-32. DOI: 10.1023/a:1010933404324
- Brown S., 2002. Measuring carbon in forests: current status and future challenges. *Environmental Pollution* 116: 363-372. DOI: 10.1016/s0269-7491(01)00212-3
- Chi H., Sun G.Q., Huang J.L., Guo Z.F., Ni W.J., Fu A.M., 2015. National forest aboveground biomass mapping from ICESat/GLAS data and MODIS imagery in China. *Remote Sensing* 7: 5534-5564. DOI: 10.3390/rs70505534
- Degnan J.J., 2002. Photon-counting multikilohertz microlaser altimeters for airborne and spaceborne topographic measurements. *Journal of Geodynamics* 34: 503-549. DOI: 10.1016/S0264-3707(02)00045-5
- Deo R.K., Domke G.M., Russell M.B., Woodall C.W., Andersen H.E., 2018. Evaluating the influence of spatial resolution of Landsat predictors on the accuracy of biomass models for large-area estimation across the eastern USA. *Environmental Research Letters* 13(5): 1-9. DOI: 10.1088/1748-9326/aabcd5
- Duncanson L.I., Niemann K.O., Wulder M.A., 2010. Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sensing of Environment* 114: 138-154 DOI:10.1016/j.rse.2009.08.018
- Freeman E.A., Frescino T.S., Moisen G.G., 2018. ModelMap: an R Package for Model Creation and Map Production. Web: <https://cran.r-project.org/web/packages/ModelMap/vignettes/VModelMap.pdf> Accessed: 10.05.18.
- Freeman E.A., Moisen G.G., Coulston J.W., Wilson B.T., 2016. Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Canadian Journal of Forest Research* 46: 323-339. DOI: 10.1139/cjfr-2014-0562
- Glenn N.F., Neuenschwander A., Vierling L.A., Spaete L., Li A.H., Shinneman D.J., Pilliod D.S., Arkle R.S., McIlroy S.K., 2016. Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. *Remote Sensing of Environment* 185: 233-242. DOI: 10.1016/j.rse.2016.02.039
- Goetz S., Dubayah R., 2011. Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Man-*

- agement 2: 231-244. DOI: 10.4155/cmt.11.18
- Gwenzi D., Lefsky, M.A., 2014. Prospects of photon counting lidar for savanna ecosystem structural studies. In ISPRS Technical Commission I Symposium, 17-20th November, Denver, Colorado, USA. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-1, pp. 141-147. DOI: 10.5194/isprsarchives-XL-1-141-2014
- Gwenzi D., Lefsky M.A., Suchdeo V.P., Harding D.J., 2016. Prospects of the ICESat-2 laser altimetry mission for savanna ecosystem structural studies based on airborne simulation data. *ISPRS Journal of Photogrammetry and Remote Sensing* 118: 68-82. DOI:10.1016/j.isprsjprs.2016.04.009
- Harding D.J., Carabajal C.C., 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters* 32: 1-4. DOI: 10.1029/2005gl023471
- Homer C., Dewitz J., Yang L.M., Jin S., Danielson P., Xian G., Coulston J., Herold N., Wickham J., Megown K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States - Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing* 81: 345-354. DOI: 10.14358/pers.81.5.345
- Houghton R.A., 2007. Balancing the global carbon budget. *Annual Review of Earth and Planetary Sciences* 35(1): 313-347. DOI: 10.1146/annurev.earth.35.031306.140057
- Houghton R.A., Butman D., Bunn A.G., Krankina O.N., Schlesinger P., Stone T.A., 2007. Mapping Russian forest biomass with data from satellites and forest inventories. *Environmental Research Letters* 2: 7pp. DOI: 10.1088/1748-9326/2/4/045032
- Hu T., Su Y., Xue B., Liu J., Zhao X., Fang J., Guo Q., 2016. Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. *Remote Sensing* 8(7): 565. DOI: 10.3390/rs8070565
- Hudak A.T., Lefsky M.A., Cohen W.B., Berterretche M., 2002. Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height. *Remote Sensing of Environment* 82: 397-416, DOI: 10.1016/s0034-4257(02)00056-1
- Jenkins J.C., Chojnacky D.C., Heath L.S., Birdsey R.A., 2003. National-scale biomass estimators for United States tree species. *Forest Science* 49: 12-35.
- Le Toan T., Quegan S., Davidson M.W.J., Balzter H., Pailou P., Papathanassiou K., Plummer S., Rocca F., Saatchi S., Shugart H., Ulander L. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment* 115: 2850-2860. DOI: 10.1016/j.rse.2011.03.020
- Lefsky M.A., Harding D.J., Keller M., Cohen W.B., Carabajal C.C., Del Bom Espirito-Santo F., Hunter M.O., de Oliveira R., 2005. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters* 32: 1-4. DOI: 10.1029/2005gl023971
- Lefsky M.A., Keller M., Pang Y., de Camargo P.B., Hunter M.O. 2007. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing* 1: 1-18. DOI: 10.1117/1.2795724
- Leigh H.W., Magruder L.A., Carabajal C.C., Saba J.L., McGarry J.F., 2015. Development of onboard digital elevation and relief databases for ICESat-2. *IEEE Transactions on Geoscience and Remote Sensing* 53(4): 2011-2020. DOI: 10.1109/tgrs.2014.2352277
- Liaw A., Wiener M., 2002. Classification and Regression by randomForest. *R News* 2/3: 18-22.
- Markus, T., Neumann T., Martino A., Abdalati W., Brunt K., Csatho B., Farrell S., Fricker H., Gardner A., Harding D., Jasinski M., Kwok R., Magruder L., Lubin D., Luthcke S., Morison J., Nelson R., Neuenschwander A., Palm S., Popescu S., Shum C.K., Schutz B.E., Smith B., Yang Y., Zwally J., 2017. The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote Sensing of Environment* 190: 260-273. DOI: 10.1016/j.rse.2016.12.029
- Marselis S., Armston J., Dubayah R., 2016. Summary of the second GEDI science team meeting. *The Earth Observer* 28(6): 31-36.
- Martino A., 2010. ATLAS Performance Spreadsheet. Web: http://icesat.gsfc.nasa.gov/icesat2/data/sigma/sigma_data.php. Accessed: 14.02.17.
- Montesano P.M., Rosette J., Sun G., North P., Nelson R.F., Dubayah R.O., Ranson K.J., Kharuk V. 2015. The uncertainty of biomass estimates from modeled ICESat-2 returns across a boreal forest gradient. *Remote Sensing of Environment* 158: 95-109. DOI: 10.1016/j.rse.2014.10.029
- MRLC, 2018. Multi-resolution land characteristics consortium (MRLC): national land cover database. Web: <http://www.mrlc.gov/> Accessed: 14.02.19.
- Narine, L.L., Popescu S., Neuenschwander A., Zhou T., Srinivasan S., Harbeck K., 2019. Estimating Aboveground Biomass and Forest Canopy Cover with Simulated ICESat-2 Data. *Remote Sensing of Environment* 224: 1-11. DOI: 10.1016/j.rse.2019.01.037
- National Aeronautics and Space Administration, 2019. GEDI - Global Ecosystem Dynamics Investigation Lidar. Web: <http://science.nasa.gov/missions/gedi/> Accessed: 11.10.18.
- National Aeronautics and Space Administration, 2017. ICESat & ICESat-2. Web: <https://science.nasa.gov/> Accessed: 10.05.18.
- Nelson R., Margolis H., Montesano P., Sun G., Cook B., Corp L., Andersen H.-E., deJong B., Pellat F.P., Fickel T., Kauffman J., Prisley S., 2017. Lidar-based estimates of aboveground biomass in the continental US and Mexico using ground, airborne, and satellite observations. *Remote Sensing of Environment* 188: 127-140. DOI: 10.1016/j.rse.2016.10.038
- Neuenschwander A.L., Magruder L.A., 2016. The potential impact of vertical sampling uncertainty on ICESat-2/ATLAS terrain and canopy height retrievals

- for multiple ecosystems. *Remote Sensing* 8(12): 1-16. DOI: 10.3390/rs8121039
- Neuenschwander A., Popescu S., Nelson R., Harding D., Pitts K., Pederson D., Sheridan R., 2017. Ice, Cloud, and Land Elevation Satellite (ICESat-2) Algorithm Theoretical Basis Document (ATBD) for Land - Vegetation Along-track products (ATL08), 78 p.
- Popescu S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy* 31: 646-655. DOI: 10.1016/j.biombioe.2007.06.022
- Popescu S.C., Wynne R.H., 2004. Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering and Remote Sensing* 70: 589-604. DOI: 10.14358/PERS.70.5.589
- Popescu S.C., Wynne R.H., Nelson R.F., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing* 29: 564-577. DOI: 10.5589/m03-027
- Popescu S.C., Zhou T., Nelson R., Neuenschwander A., Sheridan R., Narine L., Walsh K.M., 2018. Photon counting LiDAR: An adaptive ground and canopy height retrieval algorithm for ICESat-2 data. *Remote Sensing of Environment* 208: 154-170. DOI:10.1016/j.rse.2018.02.019
- Schutz B.E., Zwally H.J., Shuman C.A., Hancock D., DiMarzio J.P., 2005. Overview of the ICESat mission. *Geophysical Research Letters* 32(4). DOI: 10.1029/2005gl024009
- Simard M., Pinto N., Fisher J.B., Baccini A., 2011. Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research-Biogeosciences*, 116: 1- 12. DOI: 10.1029/2011JG001708
- Texas A&M Forest Service, 2019. East Texas Forestlands, 2012. Web: <http://texasforests.tamu.edu/fia/publications/> Accessed: 11.10.18.
- USDA Forest Service, 2018. Sam Houston National Forest. Web: https://www.fs.usda.gov/detail/texas/about-forest/districts/?cid=fswwdev3_008443 Accessed: 06.09.18.
- Zald H.S.J., Wulder M.A., White J.C., Hilker T., Hermosilla T., Hobart G.W., Coops N.C., 2016. Integrating Landsat pixel composites and change metrics with lidar plots to predictively map forest structure and aboveground biomass in Saskatchewan, Canada. *Remote Sensing of Environment* 176: 188-201. DOI: 10.1016/j.rse.2016.01.015
- Zwally H.J., Schutz B., Abdalati W., Abshire J., Bentley C., Brenner A., Bufton J., Dezio J., Hancock D., Harding D., Herring T., Minster B., Quinn K., Palm S., Spin-hirne J., Thomas R., 2002. ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics* 34(3-4): 405-445. DOI: 10.1016/S0264-3707(02)00042-X