The Development and Evaluation of a High-Resolution Above Ground Biomass Product for the Commonwealth of Puerto Rico (2000)

John S. Iiames, Joseph B. Riegel, Kristin M. Foley, and Ross S. Lunetta

Abstract

This study details the development of a U.S. Commonwealth of Puerto Rico above-ground forest biomass (AGB) product (baseline 2000) developed by the United States Environmental Protection Agency (EPA) that was compared to another AGB product developed by the U.S. Forest Service (USFS) for the same area. The USEPA product tended to over-predict in areas of low biomass and under-predict in high biomass areas when compared to observed plot data, but compared favorably to a Forest Inventory Analysis (FIA) assessment of structure and condition of Puerto Rico forests (72.6 Mg/ha versus 80.0 Mg/ha, respectively). AGB estimates were highly correlated with reference FIA biomass for both maps at their native spatial resolutions (USEPA: r =0.93, USFS: r = 0.92). AGB mean difference between both products was 33.5 Mg/ha (USFS mean = 106.1 Mg/ha; USEPA mean = 72.6 Mg/ha), a difference not out-of-scope when compared to other biomass comparative studies.

Introduction

The dependence on coastal resources (i.e., fisheries, tourism, and pharmaceutical products from natural marine resources) on the island of Puerto Rico (PR), has been threatened by municipal and agricultural growth (Abuyuan, 1999). This growth has led to declining quality and availability of drinking water and increased sediment and nutrient runoff that adversely affects coastal seagrasses, mangroves, and coral reefs (Warne *et al.*, 2005). The United States Environmental Protection Agency (USEPA) and University of South Florida are exploring the relationship between landscape composition and near-shore turbidity and chlorophyll-a concentrations across southern PR. Both turbidity and chlorophyll-*a* estimates were developed from the Moderate Resolution Imaging Spectroradiometer (MODIS) imagery data (Aqua) for 2002 to 2015. The research approach included both near-term spatially explicit (watershed-based) analysis for agricultural, urban and rural watersheds; and long-term temporal correlation analysis for individual watersheds (2001 to 2015). Landscape composition correlates included numerous landscape metrics selected to represent watershed erosion and nutrient loading potential.

One metric, above-ground biomass (AGB), allows for the assessment of the mitigating effect on the reduction of nutrient and sediment loadings during flood pulse disturbances (Bayley and Guimond, 2008). AGB estimates developed under this research will provide landscape metrics at the watershed, sub-watershed, and riparian buffer analysis scales. Here we document the construction of the USEPA AGB product (2000), then compare this product to an AGB product created by the United States Forest Service (USFS) for the Commonwealth of PR (2000). Both the USEPA and USFS biomass products incorporated similar data fusion techniques regressing satellite and/or satellite-derived (i.e. 'estimated') data with *in situ* AGB estimates. The techniques were dissimilar with respect to the estimation methods implemented and the primary predictor variables used. The USFS method utilized meteorological data and MODIS optical data as the primary predictor variables whereas the USEPA method incorporated shuttle radar data. Also, the spatial scale of the input variables and output predictions (15 versus 250 m) differed between both datasets. The requirement for a finer resolution USEPA AGB product was predicated on the nutrient flow modeling of sub-watersheds and riparian buffers, typically narrow in width to exclude the use of the coarse 250 m AGB data. The quantitative comparisons described in this paper only provide descriptors of differences and do not offer accuracy validation for either of the two products.

Numerous analysis techniques have been used to map AGB on regional and global scales (Brown *et al.*, 1989; Lefsky *et al.*, 2002 and 2005; Basuki *et al.*, 2009). Mapping efforts at these spatial scales primarily involved correlating remote sensing data products and ancillary data, with *in situ* above-ground biomass data (Zolkos *et al.*, 2013; Marks *et al.*, 2014). Each data type exhibits strengths and weaknesses in estimating information about the vertical and horizontal structure existent within a forested area (Cartus *et al.*, 2014). For example, Lefsky *et al.* (2002) and Treuhaft *et al.* (2009) mapped AGB using multi-source lidar and radar taking advantage of these sensors ability to estimate vertical tree canopy structure. Discrepancies in AGB mapping can be attributed to a number of factors including: (a) use of differing allometric equations to acquire reference AGB (Baccini and Asner, 2013); (b) differing scales and types of geospatial predictor variables (Blackard *et al.*, 2008); and (c) disagreement of predicted land-cover proportions from one technique to the other (Pan *et al.*, 2011; Harris *et al.*, 2012). Houghton *et al.* (2001) found that different AGB estimation techniques can produce AGB estimates that differ by more than two fold within the same region. Understanding the error components within each process is necessary when comparing AGB products. For example, Asner *et al.* (2011) noted that *in situ* carbon estimation alone explained 20 to 30 percent of the error in the estimation of AGB in Hawaii.

The objectives of this study were to (a) first develop a USEPA model utilizing shuttle radar data for predicting a

John S. Iiames, Kristin M. Foley, and Ross S. Lunetta are with the U.S. Environmental Protection Agency, National Exposure Research Laboratory, 109 T.W. Alexander Dr., Research Triangle Park, North Carolina 27711 (iiames.john@epa.gov).

Joseph B. Riegel is an Independent Contractor to U.S. Environmental Protection Agency, National Exposure Research Laboratory, 109 T.W. Alexander Dr., Research Triangle Park, North Carolina 27711.

Photogrammetric Engineering & Remote Sensing
Vol. 83, No. 4, April 2017, pp. 293–306.
0099-1112/17/293–306

doi: 10.14358/PERS.83.4.293

high resolution AGB product for the Commonwealth of PR, and (b) then compare this AGB product to the only other existing AGB product for this region developed by the USFS. The USEPA estimated above-ground forest biomass at a 15 m spatial resolution implementing methodology first posited by the Woods Hole Research Center where they developed the National Biomass and Carbon Dataset (NBCD2000), i.e., an above-ground forest biomass map (30 m) for the conterminous United States. For EPA's effort, spatial predictor layers for AGB estimation included derived products from the United States Geologic Survey (USGS) National Land Cover Dataset 2001 (NLCD) cover type and tree canopy density data, the USGS Gap Analysis Program (GAP) forest type classification data, USGS National Elevation Dataset (NED) topographic data, and the National Aeronautical and Space Administration's (NASA) Shuttle Radar Topography Mission (SRTM) tree height data. In contrast, the USFS biomass product (250 m spatial resolution) integrated United States Department of Agriculture (USDA) Forest Inventory Analysis (FIA) ground-based data with a suite of geospatial predictor variables including: (a) MODIS-derived image composites and percent tree cover; (b) NLCD cover type proportions; (c) NED topographic variables; (d) monthly and annual climate parameters; and (e) other ancillary variables (Blackard *et al*., 2008). Correlations (R^2) and average differences between both data sets were used to summarize the level of agreement between the two products.

Site Description

The island of PR is part of the Greater Antilles, located approximately 1,700 km southeast of Miami, Florida at approximately 18°N, 66°W. Totaling approximately 9,000 km^2 in area, the commonwealth consists of the one main island, eastern islands of Vieques and Culebra and uninhabited western islands of Mona and Desecheo (Figure 1). This island complex, formed through volcanic and sedimentary processes, is predominantly mountainous (approximately two-thirds of the island), with steeper grades on the less weathered south facing slopes. Approximately half of the island is forested (52.4 percent) with the majority (82.3 percent of the forested area) located within the ecological zone defined as moist-wet (Walker *et al*., 2007; Gould *et al*. 2008; Smith, 2002). Moisture and elevation gradients drive the diverse forest composition on the island, typically differing over short distances. Average temperature is 26°C with the northern and southern coasts receiving 1,550 and 910 mm of annual precipitation, respectively; while the interior mountains receive 3,000 to 4,000 mm annually (Larsen and Simon, 1993). The island has recovered from a largely denuded forested state (6 percent forest cover) in the late 1930's to its present state (Murphy *et al*., 2012). Over 500 native tree species are found throughout the island where most forest composition consists of secondary forest, with only 5.2 percent being protected.

Methods

The USEPA high resolution AGB map was developed at a 15 m spatial resolution, then aggregated to 250 m (simple averaging) to spatially align with that of the USFS AGB map for the same baseline year (2000) to support a comparison of AGB comparison across the main island of PR. The eastern islands of Vieques and Culebra were excluded due to the area extent of the USFS effort. The USEPA AGB methodology incorporated regression tree-based algorithms to identify the most

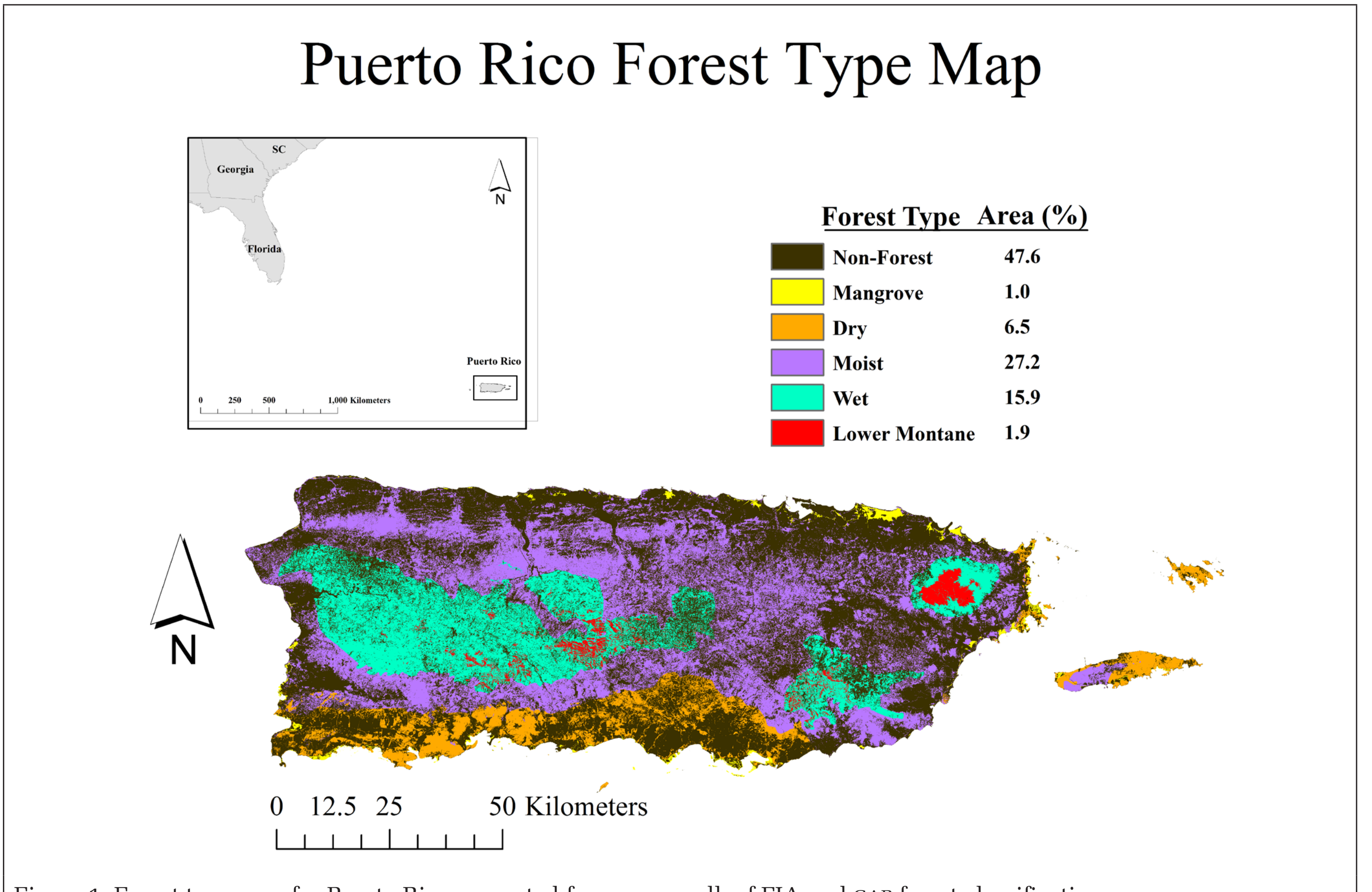


Figure 1. Forest type map for Puerto Rico generated from crosswalk of FIA and GAP forest classifications.

important variables in the multivariate analysis. This product was then compared to the USFS AGB product to provide a baseline comparison of how the new source data impacts the estimation of AGB. Further details on the development of the USFS product can be found in Blackard *et al.*, 2008.

USEPA AGB

The USEPA AGB product was created using multivariate regression tree analysis. This effort included modeling a suite of wall-to-wall raster geospatial predictor variables with biophysical response variables measured on FIA plots located throughout the island. A forest/non-forest mask was created from the GAP dataset for PR where five major forest types were identified for forest biomass estimation (Gould *et al.*, 2008). Data used to support the analysis included optical (Landsat ETM+), digital orthorectified quarter quadrangles (DOQQ), and Radar Interferometric Synthetic Aperture Radar (INSAR). These data were used to produce the ancillary predictor variable data layers used in this regression tree analysis.

Modeling Data

As per the previous conterminous United States AGB mapping by Walker *et al.* (2007), we compiled both predictor (remotely sensed and ancillary) and response (FIA generated) variables, required for multivariate tree-based regression models where observed canopy height and AGB are related to remotely sensed and ancillary data sources. The primary datasets for deriving the predictor variables included: (a) canopy density from the NLCD2001; (b) digital elevation from the NED; (c) canopy height from the 2000 SRTM; and (d) 2001 land-cover (LC) from GAP. These data types were reprojected from their native formats and resampled (bilinear interpolation) to match the PR GAP-LC product in a State Plane projection (15 m) horizontally referenced to the NAD83 Datum.

Canopy Density

The NLCD2001 Percent Tree Canopy product (CanDen) depicts the spatial distribution of tree canopy density for trees ≥5.0 m tall as a continuous variable with values ranging from 1 to 100 percent (Huang *et al.*, 2001). The NLCD percent canopy data (30 m), were downloaded from *http://seamless.usgs.gov* in the native Albers Equal Area Conic projection horizontally and vertically referenced to NAD83 Datum and the GRS80 Ellipsoid, respectively, then reprojected (State Plane, NAD83).

NED Data

The NED digital elevation map (DEM) data with a 30 m spatial resolution, were downloaded from *http://seamless.usgs.gov*. These data were registered in the native geographic (unprojected) coordinate system horizontally and vertically referenced to the NAD83 Datum and the North American Vertical Datum of 1988, respectively. The data were first mosaicked and clipped to the region of interest, then reprojected (State Plane, NAD83). The NED 30 m DEM's were created using 1:24 000-scale cartographic contours, generally of lower overall quality than other sources derived from lidar or digital photogrammetry. All these data were derived from aerial photography collected from 1941 to 1972. USGS DEM data quality objectives were to achieve a vertical accuracy RMSE of ≤7.0 m, with a maximum allowed RMSE of 15 m. Other elements that affect DEM accuracy were the horizontal grid spacing, the collection and processing procedures, and digitizing.

SRTM Data

The SRTM employed an Interferometric Synthetic Aperture Radar to collect both X-band (3.1 cm, 9.6 GHz), and C-band (5.6 cm, 5.3 GHz) data (11-22 February, 2000) at a swath width of 225 km. The C-band data was used to derive vegetative height data providing near complete coverage in the 56°S to 60°N latitude range (Hoffman and Walter, 2006). Mission expectations for absolute horizontal and vertical accuracies were 20 and 16 m, respectively. However, absolute vertical error was shown to be much smaller (~5.0 m), based on independent evaluations by the USGS National Imagery and Mapping Agency (NIMA) and the SRTM project team (Curkendall *et al.*, 2003; Rosen *et al.*, 2001a,b; Smith and Sandwell, 2003).

Puerto Rico was covered by two SRTM passes at look angle/direction for pass one: 53°/58° and pass two: 50°/-59°. SRTM non-void filled (available data) elevation data, processed by NASA's Jet Propulsion Laboratory (JPL), was downloaded from the USGS EarthExplorer website (*http://earthexplorer.usgs.gov*) in 1-arc sec (~30 m) band interleaved (BIL) tiles. Eight tiles ranging in latitude (N 17° to 19°) and longitude (W 69° to 65°) were mosaicked in the native unprojected geographic coordinate system horizontally and vertically referenced to WGS84 Datum and the EGM96 Geoid, respectively. SRTM height data was registered to the mean scattering phase center (MSPC) within a vegetated canopy located below mean and maximum canopy height (Simard *et al.*, 2006). The MSPC was dependent on target and sensor characteristics including vegetation structure and moisture, soil roughness and moisture, wavelength, baseline length and attitude, polarization, incidence angle, and phase noise. For example, the C-band scattering mechanism for vegetated canopies was largely volumetric, where depth of radar penetration is dependent on the canopy density (Sarabandi and Lin, 2000). Of all the potential error sources in estimating MSPC, the largest source was the phase noise, which is caused by thermal and quantization noise of the radar receivers (Bamler, 1999).

GAP-LC

The PR GAP-LC dataset, was selected over the NLCD2001 to provide finer spatial resolution (15 m versus 30 m). It was derived from Landsat-7 ETM+ imagery and ancillary data (i.e., climate, substrate, topography, and disturbance) collected between 1999 to 2003 (Gould *et al.* 2008). Multiple image dates were acquired over the four World Reference System 2 (WRS2) path/rows to ensure cloud-free data for most regions. This 15 m product was projected into the State Plane projection and horizontally referenced to the NAD83 Datum. The GAP-LC included 70 cover classes based on a hierarchical classification scheme produced an overall accuracy of 84.9 percent and a Kappa statistic of 0.80. User's and Producer's accuracies were greater than 80 percent for all major classes (Forest, Mangrove, Grassland-Pasture-Agriculture, Urban-Barren, Water) with the exception of the Woodland-Shrubland class (52 to 72 percent). Walker *et al.* (2007) proposed a cross-walk product using the GAP-LC classes with the corresponding FIA forest type classes. The resulting product contained five FIA forest LC classes: (1) mangrove (1.0 percent area), (2) dry forest (6.5 percent area), (3) moist forest (27.2 percent area), (4) wet and rain forest (15.9 percent area), and (5) lower montane wet and rain forest (1.9 percent area). This cross-walked map served as the spatial cover type extent for the USEPA AGB map.

FIA Data

Biomass and canopy height data were extracted from the FIA database that included 397 unique plots surveyed at least once (some twice) between 2001 to 2009 and which had been listed as 'forested' by at least one survey. Only plots sampled within five years of the year 2000 timeframe were included (2001 to 2005). Field plot design consisted of four 7.3 m fixed radius plots for the collection of the following data: (a) forest type, (b) site attributes, (c) species, (d) height and diameter, and (e) condition. Included in the database were AGB estimates per tree as calculated from predetermined allometric equations per forest type (Table 1). The FIA definition of live AGB included "in live tree bole wood, stumps, branches and twigs for trees 2.54 cm diameter or larger and is derived from region- or species-specific allometric equations." (Blackard *et al*., 2008). Following Walker *et al*. (2007), the following plot-level statistics were calculated for each plot: (a) AGB (Mg/ha), (b) maximum height (m), (c) average plot tree height (m), and (d) basal-area weighted average height (m) using the equation:

$$\sum_{i} \rightarrow n\ [(BA_i / BA_{plot}) * Height_i)], \quad (1)$$

where n is the number of trees, BA is the basal area, Height is tree height (m), and i is the i^{th} tree.

Image Segmentation

Image segmentation was required to deal with the issue of SRTM phase noise. Following the methodology of Kellendorf *et al*. (2006), partitioning of the landscape into units of similar height and overall structure was implemented to produce relatively uniform image objects that were of sufficient size to provide for adequate sample averaging and noise reduction in forested regions. Segments were designed to be homogenous in terms of topographic slope, vertical (canopy height) and horizontal (canopy density) forest structure, and forest stand type. Segmentation was performed using the feature extraction tool in Environment for Visualizing Images software (ENVI v. 4.8). The estimated canopy height (SRTM minus DEM) and the CanDen from the NLCD layers were used in the segmentation. Image segmentation was performed in a two-step process targeting an average polygon area of 15 to 20 pixels (1.35 to 1.8 ha), large enough to produce a reliable estimate of canopy height yet small enough to retain a uniform canopy height (Walker *et al*., 2007). The first segmentation included all the non-void filled forested pixels, including those with anomalous and non-anomalous estimated canopy height values. The second segmentation was performed on both the anomalous and non-anomalous areas separately and then were later recombined. The eight predictor variables calculated for each polygon are listed in Table 2.

Table 1. Forest Biomass Allometric Equations Used to Calculate FIA Biomass per Tree

Forest Type	Equation
Lower montane wet and rain forest	[1]$AGB = 4.7962 + 0.031 * dbh^2 Ht$
Subtropical wet and rain forest	[2]$AGB = e^{(0.95 * \ln dbh\ Ht - 3.282)}$
Subtropical moist forest	[3]$AGB = e^{(-1.71904 + 0.78214 * \ln dbh\ Ht)}$
Subtropical dry forest	[4]$AGB = e^{(-1.94371 + 0.84134 * \ln dbh\ )}$
Mangrove	[5,6]multiple

[1]Weaver and Gillespie, 1992; [2]Scatena *et al*., 1993; [3]Brandeis *et al*., 2006; [4]Brandeis *et al*., 2007; [5]Cintron and Schaeffer-Novelli, 1984; [6]Fromard *et al*., 1998

Polygon Segments and FIA Plot Intersections

Exact coordinates for FIA plot locations are not available due to the public federal law protecting the privacy of the landowner. However, with USFS assistance, FIA plot locations were spatially linked with remote sensing derived products by the Southern Research Station FIA Research Work Unit 4801 in Knoxville, Tennessee facilitating the joining the eight object-based metrics with the FIA-based plot metrics. Plots removed from analysis included plots that were only partially forested, plot areas that were not sampled due to landowner access denial or hazardous conditions, and plots that intersected segment polygons <1.5 ha. A complete list of variables (response and predictor) for this analysis is found in Table 2.

Modeling Canopy Height and AGB

SRTM vegetation canopy height required calibration with *in situ* height estimates to compensate for C-band radar penetration into the canopy (Kellndorfer *et al*., 2004; Dubayah *et al*., 2007; Sexton *et al*., 2009). Following the segmentation and intersections, canopy height was predicted by modeling the relationship between the eight predictor variables and the two FIA height response variables (MaxHt and BAWAveHt). The predictor variables AveCanHt and SdCanHt (from SRTM and DEM data), calculated at the polygon segment level, were converted to a 15 m raster format, where each cell within the polygon of origin retained the same value. Following Walker *et al*., (2007), the original (not averaged) elevation, canopy density, and slope raster datasets were used instead of using the AveElev, AveCanDen, and AveSlope values from the segmented polygons.

The decision tree classifier algorithm randomForest (RF), run in R (Maechler *et al*., 2013), was executed to determine the Variables of Importance Plots (VIP) for predicting canopy height. The RF algorithm (Breiman, 2001) tested a large collection of decorrelated decision trees that produce a vote for the most popular class. The responses generated by each tree were averaged to obtain an estimate of the dependent variable. Walker *et al*. (2007) noted that the algorithm cannot predict beyond the range in the training data and may also over-fit data (Walker *et al*., 2007). However, the RF algorithm does guard against model over-fitting limiting the need for cross-validation (Fassnacht, 2014). Biomass was modeled following the same procedures implemented in estimating canopy height, with the only difference being the ingestion MaxHt and BAWAveHt as predictor variables. AGB and canopy height maps (MaxHt and BAWAveHt) were created using the best fit model from the regression tree estimation.

Table 2. Predictor and Response Variables Implemented in USEPA Height and Biomass Modeling

Variable	Description	FIA-based reference variables	Object-based variables	Response or predictor variables
MajForType	Majority Forest Type	X		P
BAWAveHt	Basal-Area Weighted Average Height	X		R
MaxHt	Maximum Height	X		R
AveCanHt	Average Canopy Height (SRTM)		X	P
SdCanHt	S.D. Canopy Height (SRTM)		X	P
AveElev	Average Elevation (DEM)		X	P
SdElev	S.D. Elevation (DEM)		X	P
AveCanDen	Average Percent Canopy (NLCD)		X	P
SdCanDen	S.D. Percent Canopy (NLCD)		X	P
AveSlope	Average Slope (Dem)		X	P

USFS AGB

The USFS created an AGB product for the conterminous US, Alaska, and PR for year 2000 (Blackard *et al.*, 2008). The general methodology employed was similar to the techniques used within this study to create the USEPA AGB map including segmentation (at a national level) and regression tree analysis. Predictor variables included surface reflectance bands at 500 m resolution from the MODIS sensor (2001), MODIS-derived products including vegetation indices (250 m) and percent tree cover (i.e., Vegetation Continuous Fields) (500 m) (*http://glcf.umd.edu/data/vcf/*); NLCD-derived data (30 m); NED (30 m) topographic variables (i.e., slope, dominant aspect, elevation); monthly and annual climate parameters (1961 to 1990) including average monthly and annual precipitation and temperature measures (420 m); and other ancillary variables. Forested LC was estimated from a 26 class LC map developed by the USDA Forest Service (Helmer *et al.*, 2002) using Landsat imagery based on a supervised classification approach. LC proportions were summarized using a 9 × 9 moving average window then resampled to the base 250 m resolution. All other data were resampled to this same resolution (Blackard *et al.*, 2008).

Response variables, including forest condition and live AGB estimate, were extracted from the FIA database. These predictor variables were then spatially linked to the response variables. The non-parametric data mining package (See5) was used to first create a forest/non-forest binary classification. Next, the regression tree-based algorithms in Cubist were used to predict biomass. Finally, forest biomass was delineated using this forest/non-forest classification (Quinlan, 1986; Quinlan, 1994). Uncertainty maps were created for both the forest/non-forest map and the map of the above-ground biomass product. All data were downloaded from the USFS Geodata Clearinghouse (*http://data.fs.usda.gov/geodata/rastergateway/index.php*) in the Lambert Conformal Conic projection at 250 m (Datum, NAD27; Spheroid, Clarke 1866), projection parameters specific to PR.

USEPA and USFS Product Comparisons

The USEPA AGB map could not be assessed for accuracy using traditional techniques due to the lack of an available reference data to support a statistically robust validation. The only biomass reference data available for Puerto Rico were the FIA biomass plot data. However, because these data were used (needed) for biomass model development, they could not be used to perform an independent validation of model performance. The issue of available high quality reference data for model validation is a frequent problem within the scientific community. One approach commonly used is to compare the results (estimates) from multiple independent modeling efforts (Mitchell *et al.*, 2011). The greater the correspondence of modeling estimates, the greater the confidence, and vice versa.

To facilitate a comparison between AGB modeling results the USEPA product was first aggregated to the 250 m grid using equal area averaging and reprojected to match the USFS product (Lambert Conformal Conic projection; Datum NAD27; Spheroid - Clarke 1866). Modeling comparisons included assessing forested area (EPA and USFS estimates), the NLCD CanDen layer, and both AGB products.

Forest versus Non-Forest

Binary maps of forest (1) and non-forest (0) were compared between three datasets: USFS, USEPA, and the NLCD CanDen data layer. The CanDen data layer is a continuous product estimating canopy density over the 30 m pixel. Tree canopy was not constrained only to NLCD forest class pixels but also included the wetland and urban classes. CanDen pixels were classified as 'forest' if one or more 30 m CanDen pixels within a 250 m grid cell had values >0.0. Forest versus non-forest was assessed overall between the three datasets, evaluating agreement and non-agreement across the main island. The USEPA and NLCD CanDen data layers were also compared at 20 percent crown density intervals (i.e., 0 to 20 percent, 20 to 40 percent, etc.).

Above Ground Biomass

The purpose of comparing the two AGB products was only to investigate the level of agreement. A comparison of the two AGB datasets was conducted for overall absolute biomass stratified at 20 percent crown density levels (determined from the NLCD2001 Percent Tree Canopy layer) and across five FIA LC class categories. Absolute differences were also calculated along 90 km transect profiles and relative differences were computed to assess directional agreement (i.e., increases versus increases, increases versus decreases, etc.). Descriptive statistics were calculated from a simple absolute difference map. Correlation (R^2), RMSE, and bias were estimated in the comparison of the AGB products across 20 percent crown density intervals and across the five FIA forest class categories including mangrove, dry forest, moist forest, wet/rain forest, lower Montane/rain forest.

Also, four 90 km west-to-east transects (three pixels wide) were selected (non-random), equally spaced north to south, to include both the central mountain region along with the coastal northern and southern portions of the island. The predicted AGB was used to display differences for multiple forest types along differing elevation grades. AGB slope (percent change between neighboring pixels) was computed for each AGB dataset then differenced to evaluate direction and magnitude of AGB change. In general, AGB comparisons can serve to elucidate uncertainty in estimates of carbon cycling. Zolkos *et al.*, 2012, found AGB accuracy issues were attributed to systematic differences in sensor type, forest biome, and plot sizes used for *in situ* calibration and assessment. This comparative study also probes the differences between the two AGB products, not looking to ascribe one versus the other as 'more correct.'

Results

Independent Variables

Understanding input data inherent error and how it might affect AGB estimates is necessary to identify possible issues with overall model integrity. Erroneous, imprecise, and inaccurate data can 'propagate' into new data layers and may either be additive or multiplicative, thus proving very difficult to predict (Lunetta *et al.*, 1991; Iiames *et al.*, 2015). The PR GAP-LC dataset used to produce the forest/non-forest mask for the USEPA AGB product was mapped with high accuracy for the 'forest' cover type showing user and producer accuracies of 86.4 percent and 91.5 percent, respectively (Kappa = 0.80). However, the 'woodland/shrubland' class was mapped with more uncertainty (user/producer = 72.1 percent/52.5 percent; Kappa = 0.67).

Regarding the CanDen dataset, a pilot-based study that applied regression tree techniques revealed a mean absolute difference between actual and predicted canopy density of 9 to 11 percent on three sites within the continental United States (CONUS) (Huang *et al.*, 2001). In another study, forest tree canopy cover was significantly underestimated 11.7 percent (SE = 1.4 percent) in 57 of 66 NLCD mapping zones within the CONUS (Nowak and Greenfield, 2010). The layer had previously shown limitations associated with high canopy closure forests. Cartus et al. (2014) noted that the CanDen predicted AGB best in the early successional stage of forest development, with densities below 100 percent. Interestingly, over 33 percent of the forested biomass mapped for this study resided within the 80 to 100 percent tree canopy interval, which is less than optimal for biomass prediction.

SRTM data voids, less than 0.02 percent (0.78 km^2) of the forested area, were identified within the USGS GAP-FIA crosswalk dataset forest/non-forest extent. Inherent uncertainty within both the SRTM MSPC and the DEM datasets confounded tree height values produced in the differencing of these two datasets. SRTM height data was registered to the MSPC within a vegetated canopy located below mean and maximum canopy height. The largest potential error source in estimating MSPC was phase noise; however, this was mitigated to some extent through the segmentation process.

DEM vertical accuracy varies significantly because of differences in source data quality, terrain relief, LC, and other factors. RMSE for CONUS-assessed DEMs showed a difference of 0.84 m between ground-view obstructed forest cover (evergreen) and non-obstructed forest cover (deciduous) (Gesch *et al.*, 2014). For this study, source data quality was also an issue as observed by comparing DEMs from 2003 (RMSE = 1.55) and 2013 (RMSE = 2.44). Within our study, the map of estimated canopy height (SRTM minus DEM) revealed areas where the estimated canopy height seemed unrealistically low ($\leq$0 m) or high (>50 m). Approximately 25.2 percent of the non-void forest pixels fell into the first category, 0.02 percent into the latter (Table 3), and 95.0 percent of all negative values fell within −1 m to −9 m (Figure 2). These errors can be attributable to inherent SRTM noise where the USGS found absolute vertical errors approximately 5.0 m. Figure 3 documents one example of a south-facing slope with an abundance of height values between −1 m and −9 m. A comparison to 2004 imagery within this figure shows that the negative height area corresponds to areas of fractional tree cover, dominated by shrub and grass cover types. An inspection of extreme negative values ($\leq$−20 m) revealed that these pixels were not randomly distributed over the landscape; instead, there were obvious spatial patterns in their arrangement (Figure 3).

In some cases, patches of high negative values were observed on south-facing slopes suggesting that these errors may be attributable to errors in the SRTM data. In other cases, the negative values seem to become more extreme along an elevation gradient (i.e., less extreme negative values on lower elevations versus more extreme negative values at higher elevations). These anomalous areas were likely attributable to errors in the DEM data rather than random noise in the SRTM data. The NED was based on outdated (1941) source imagery for one large area with extreme negative values. The highest percentage of negative canopy heights were located in the Lower Montane FIA forest (23.4 percent), with fairly even

Table 3. The Distribution of Tree Heights Derived from the STRM and DEM Height Data Sets (SRTM minus DEM)

Type	Area (km2)	Area (%)
Anomalies - Low (CanHt < 0 m)	2256	25.2
Anomalies - High (CanHt > 50 m)	2	< 0.1
Non Anomalies	6631	74.4
SRTM Voids	2	< 0.1
Total	8892	100.00

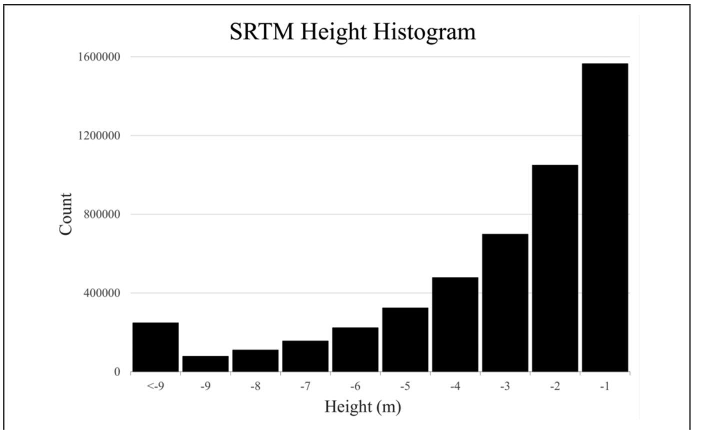


Figure 2. SRTM height - DEM histogram of negative height values.

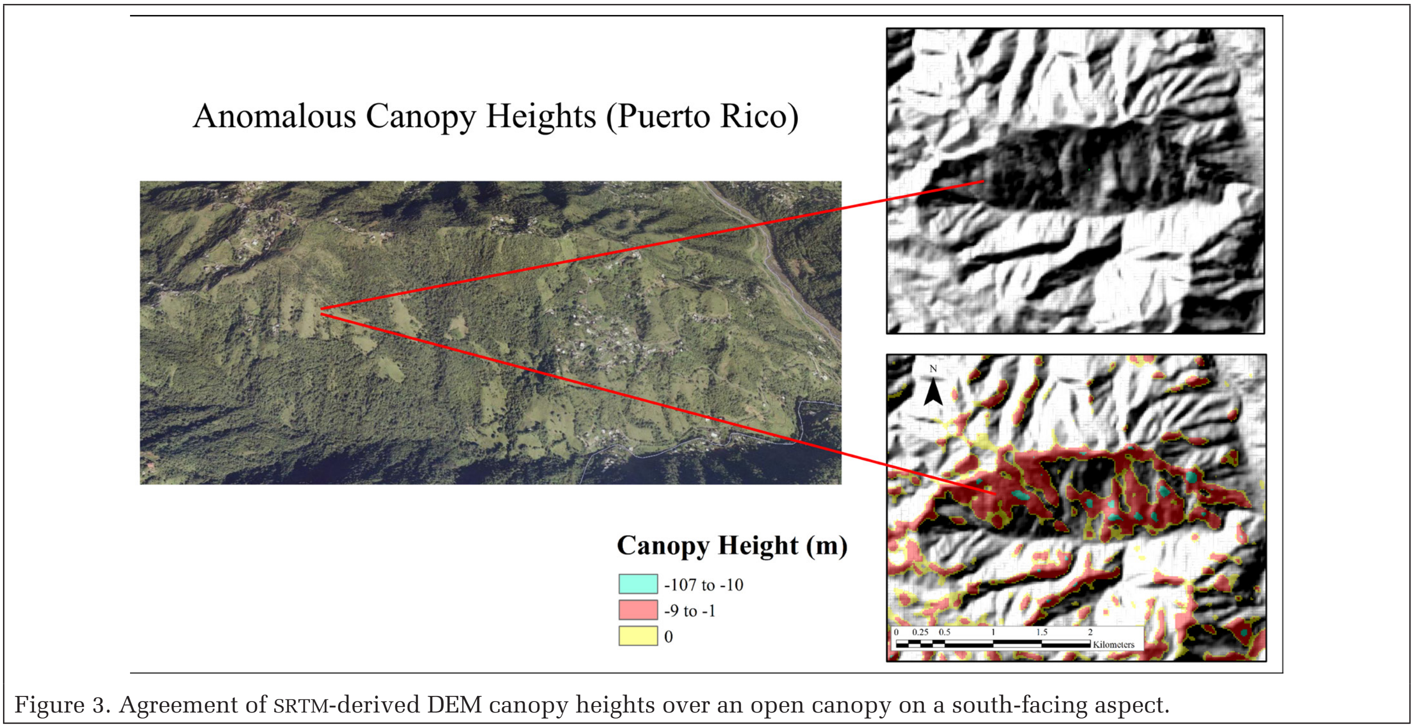


Figure 3. Agreement of SRTM-derived DEM canopy heights over an open canopy on a south-facing aspect.

distribution (10.7 to 16.0 percent) across the other four FIA forest classes.

There were fewer pixels with unusually high estimated canopy height values. In one case, it was clear that these high anomalous values were due to the construction of a dam (evident in high-resolution imagery) between the data of the DEM source imagery and date of the SRTM data collection. Also, a different location indicated the existence of a low valley while the SRTM data depicted a reservoir; also likely a date mismatch. For the remaining areas with unrealistically high canopy height values, there were no obvious patterns in their arrangement or obvious explanations for the high values. These anomalies were most likely due to noise in the SRTM (Hoffmann and Walter, 2006).

Segmentation and FIA Plot Intersections

The segmentation procedure produced average polygon sizes of 0.18 to 0.66 ha (Table 4), less than the targeted average polygon area of 1.35 to 1.8 ha (Walker *et al.*, 2007). These smaller polygons were the result of the fragmented non-void forest SRTM data that reduced the average polygon size (the correction procedure further fragmented the polygons). The trade-off in adjusting the input parameters to produce larger average polygon sizes had the negative effect of also increasing the maximum polygon size. Therefore, we retained the input parameters to provide a balance between a relatively large average polygon size and relatively small maximum polygon size.

Only 223 of 397 unique forested plots in the FIA database, intersected with polygons constrained to the GAP forest cover type. This disparity was likely attributable to the definitions used by the FIA and GAP programs. The FIA regarded areas ≥10 percent canopy coverage as 'forested', whereas the GAP defined ≥25 percent canopy cover as 'woodland' and ≥60 percent as 'forested' cover types. This more liberal forested definition used by the FIA, can be seen in the estimated forested land for both datasets: FIA (57 percent) and GAP (52.4 percent) (*http://srsfi a2.fs.fed.us*). Of the 223 FIA plot intersections, 203 (91 percent) matched the FIA forest type and the majority forest type (MajForType) cross-walked to the GAP forested cover type of the associated polygon.

Next, we vetted these 223 plots for data reliability (i.e., tree measurement data present, incomplete plot sampling due to lack of access issues, etc.). A total of 20 plots were omitted because they contained incomplete tree measurement data. Walker *et al.* (2007) removed these partially sampled FIA plots to minimize introducing error into the regression-tree modeling process in order to reduce bias in the final estimation. An additional 16 plots were removed due to a sampling time difference greater than five years from the year 2000 target base-year. Finally, 33 plots were removed that intersected polygons with areas less than 1.5 ha. After the vetting process, a total of 154 plots remained for the modeling effort.

Table 4. Segmentation Unit Area Results for Puerto Rico Including the Main Island, Vieques and Culebra (Non-voids are SRTM Data with Data Values)

Segment	Pixels Used	Polygons (*n*)	Average Area (ha)	Max Area (ha)
1	Non-Void / All	707170	0.66	9586.44
2	Non-Void / Anomalies +	432540	0.18	137.77
3	Non-Void / Non-Anomalies	822547	0.47	1914.68

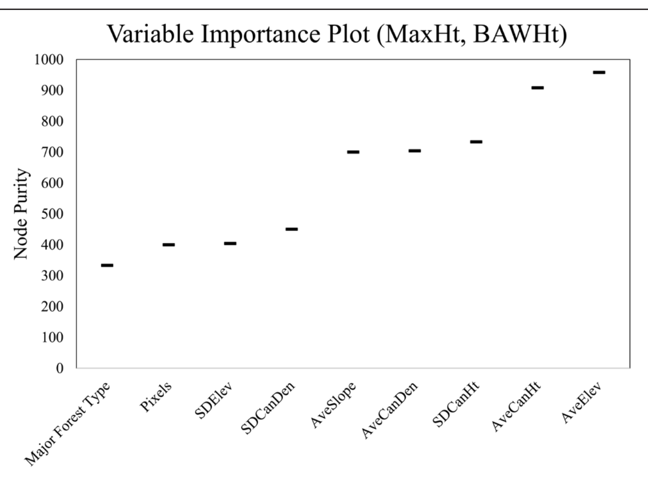


Figure 4. Variable Importance Plot (VIP) for predicting BAWAveHt and MaxHt.

Canopy Height

MaxHt and BAWAveHt were modeled using Segmentation-1 (anomalous and non-anomalous plots combined; *n*=154) and Segmentation-2 (non-anomalous only; *n* = 115) combined with anomalous plots only (*n* = 39). Model outcome was similar for both segmentations, therefore Segmentation-1 variables of importance were used in modeling these relationships (Table 2). The VIP top predictor variables for both MaxHt and BAWAveHt modeling were as follows: AveElev, AveCanHt, AveCanDen, SdCanHt, and AveSlope (Figure 4). A comparison with the Walker et al. (2007) study found that the FIA forest type variable was ranked as the highest in importance, where our modeling ranked that particular variable as least important. This may have been due to elevation gradients in PR dictating forest type change more significantly than found in the Walker *et al*. (2007) study (Brandeis, 2003).

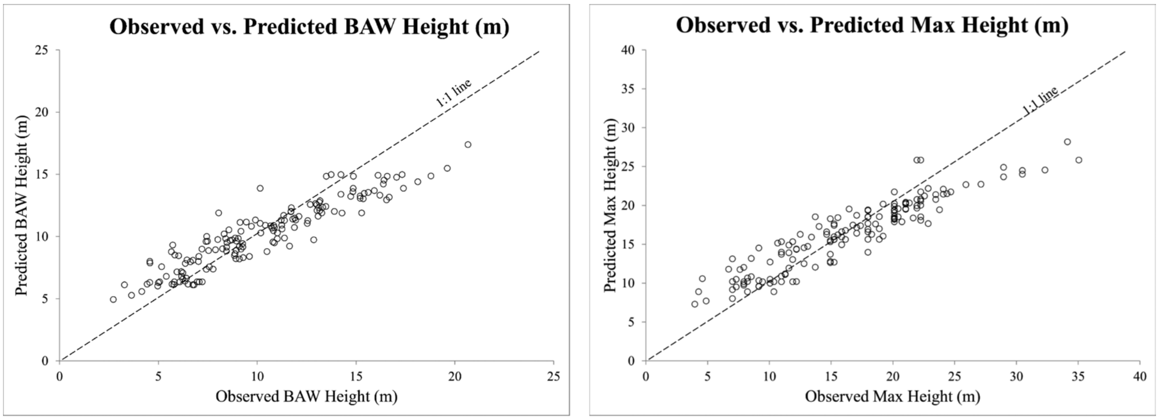
The final MaxHt (*r* = 0.93, mean error = 2.2 m) and BAWAveHt (*r* = 0.93, mean error = 1.3 m) models were roughly of similar quality and also very comparable to those produced by Walker *et al* (2007) (Table 5). Predicted versus observed plots for the MaxHt and BAWAveHt showed a tendency for the models to over-predict low canopy height values and under-predict high canopy height values (Figure 5). This response was most pronounced in areas of low forest densities and soil/canopy moisture variations. In these two scenarios, the L-band SAR backscatter intensity signal tends to saturate (Imhoff, 1995; Koch, 2010). This trend was also seen in Walker's results (Walker *et al.*, 2007). Coupled with the SRTM data issues were the inherent DEM errors. Source imagery for PR DEMs were from 1941 to 1972, 30 to 60 years removed from the base year 2000, possibly exacerbating these modeled height issues. Mean error was 26.0 percent less when comparing the height mapping results from MaxHt (*r* = 0.68, mean error = 3.9 m) versus BAWAveHt (*r* = 0.65, mean error = 2.4 m) (Table 6).

Table 5. Pixel-Based Height Model Correlations

Variable	*n*	Var (*n*)	*r*	Mean Error	MeanHt	Mean Error/ MeanHt
MaxHeight	187	5	0.93	2.2 m	16.4 m	0.1 m
BAWAveHt	187	5	0.93	1.3 m	10.2 m	0.1 m

Table 6. Polygon-Based Height Model Correlation Analysis Results

Height Stat	Polygon size range	Plots (*n*)	*r*	Mean Error (m)
MaxHt	< 1.5 ha	33	0.53	5.3
MaxHt	> 1.5 ha	154	0.72	3.6
MaxHt	All	187	0.68	3.9
BAWAveHt	< 1.5 ha	33	0.44	3.4



(A) (B)

Figure 5. Observed versus predicted BAW (A), and Max (B) forest height across FIA plots.

BAWAveHt	> 1.5 ha	154	0.72	2.2
BAWAveHt	All	187	0.65	2.4

A comparison of the polygon-based height model correlations (Table 6) and the pixel-based height map correlations (Table 5) revealed better fits with both modeled MaxHt (r = 0.93) and BAWAveHt (r = 0.93) compared to mapped MaxHt (r = 0.68) and BAWAveHt (r = 0.65). These differences may be attributable to plot size difference between the FIA plot (~9,500 m^2) and the corresponding pixel (225 m^2). Also, potential model over-fitting of the data, which makes extrapolation less reliable, especially for predicting values beyond the range of the input data.

AGB Modeling

Predictor variables used in the biomass modeling included: predicted MaxHt or BAWAveHt (from the canopy height map), CanDen, DEM, slope, and FIA forest type. The modeling procedure followed the canopy height modeling procedure, ultimately identifying five predictor variables from the RF output VIP. The FIA forest type variable was the only variable excluded. MaxHt and BAWAveHt were run separately with the remaining three predictor variables. The correlation coefficient for predicted and observed biomass was calculated for each model (Table 7). Overall, the MaxHt and BAWAveHt models were equally successful (r = 0.93). Similar to the canopy height models, the biomass models tended to underestimate high observed biomass values and overestimate low values (Figure 6).

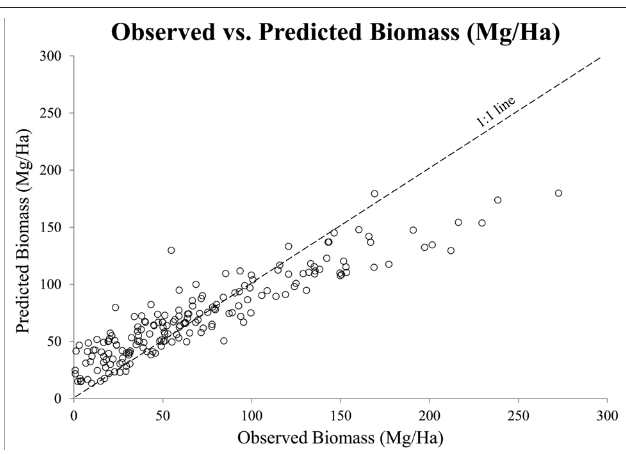


Figure 6. USEPA observed versus predicted biomass across FIA plots.

Table 7. Pixel-Based Biomass (Mg/ha) Model Correlation Results

Variable	n	Var (n)	r	AGB Error (μ)	AGB (μ)	AGB Error (μ) / AGB (μ)
MaxHeight	187	4	0.93	19.7	71.9	0.274
BAWAveHt	187	4	0.93	19.6	71.9	0.273

Table 8. USEPA-Modeled Forest Biomass (Mg/ha) for Puerto Rico

Map	Pixels (15 m)	Max	Min	Mean	SD
EPA (forest only)	20722836	247.46	10.75	72.59	26.83
EPA (all)	39520907	247.46	0	38.07	41.13

Table 9. USEPA forest/non-forest mask compared to NLCD2001 Forest/Non-Forest Mask at 20 Percent Canopy Density Intervals

Forest (%)	EPA Pixels (n)	Percent	NLCD Pixels (n)	Percent
0.1 - 20	24475	19.3	24053	20.3
20 - 40	18259	14.4	19213	16.2
40 - 60	18696	14.7	18743	15.8
60 -80	20434	16.1	17603	14.8
80 - 100	45117	35.5	39142	33.0
0.1 - 100	126981	100.0	118754	100.0

Table 10. Comparisons of 250 m Forest/Non-Forest Mask Between the Reference
Dataset (NLCD) and the USFS and USEPA Products

		TOTAL	EPA (only)	USFS (only)	NLCD (only)	Agree
250 m Pixel (n)	EPA vs NLCD	141517	13098	*	5445	122974
	EPA vs USFS	141517	47377	984	*	93156
	NLCD vs USFS	141517	*	3358	42098	96061
PERCENT	EPA vs NLCD	141517	9.3	*	3.8	86.9
	EPA vs USFS	141517	33.5	0.7	*	65.8
	NLCD vs USFS	141517	*	2.4	29.7	67.9

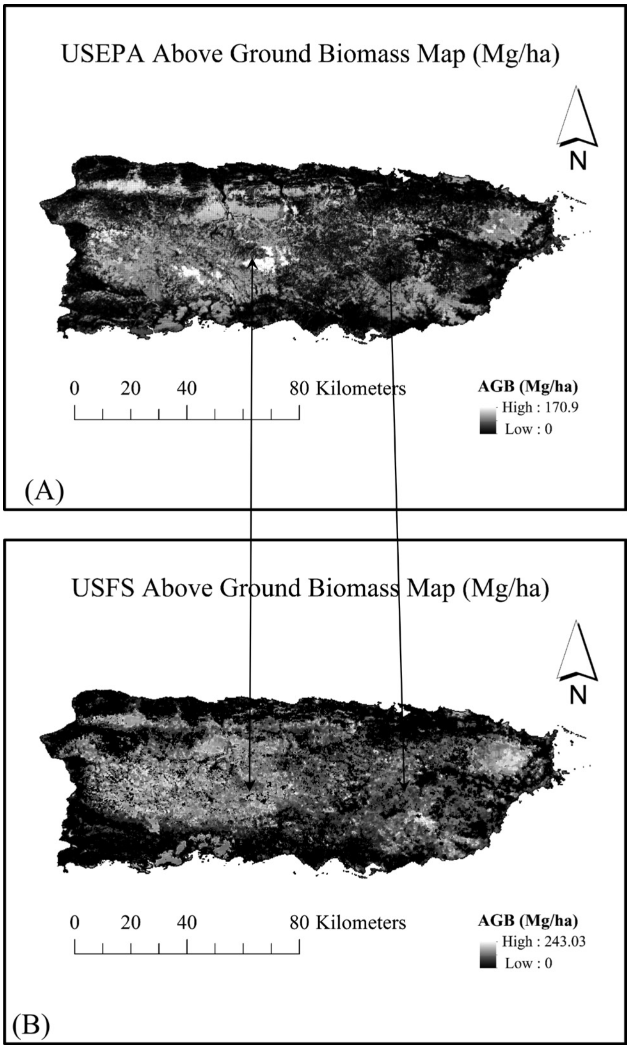


Figure 7. AGB maps (250 m) of USEPA (A), and USFS (B). Two areas indicated by arrows show high relative AGB increase and decrease, left and right arrows, respectively, for the USEPA map when contrasted with adjacent pixels. The USFS AGB map, in contrast, shows more gradual change between these two areas.

Table 11. Comparison of USFS and USEPA AGB 250 m Products (Mg/ha)

Map	Pixels (250 m)	Max	Min	Mean	SD
EPA (forest only)	126981	170.9	< 0.1	42.1	33.0
EPA (all)	141517	170.9	10.8	37.7	33.8
USFS (forest only)	80590	243.0	3.8	106.1	36.3
USFS (all)	141517	243.0	0	60.4	59.2

Finally, a AGB product was created using the MaxHt biomass model. It was created on a pixel-by-pixel basis using the same methodology used to create the canopy height maps. Mean biomass (forest only) for the 15 m pixels was 72.59 Mg/ha (σ = 26.83) (Table 8). This estimate is close in agreement to an assessment of structure and condition of PR forests (2003) (Brandeis, 2006) where mean AGB was estimated at 80 Mg/ha. This study estimated that 66 percent of the forest structure was classified predominately in the early stages of development, 18 percent as forest reversions (recently recolonized with trees), and the remainder as mature secondary forests. Over 93 percent of the forest stands were represented by the sapling-seedling stage or small diameter stage (12.5 cm – 22.4 cm dbh).

Forest Versus Non-Forest

Comparing the 250 m forest/non-forest masks for both the USEPA to the NLCD products showed that the USEPA classified approximately 6.5 percent more area as forest than the NLCD product at the 20 percent canopy density level (Table 9). Scaling the USEPA and NLCD forest/non-forest masks to 250 m allowed for comparison agreement evaluations between the NLCD, USFS, and USEPA products. The forested component was similar with the NLCD (84.3 percent) and the USEPA (89.7 percent) compared to the USFS (56.9 percent) products (Table 10). Assuming the NLCD forest/non-forest mask as the reference, the USEPA had the best agreement compared to the USFS, 86.9 percent and 67.9 percent, respectively (Table 10).

AGB Absolute Difference Comparison

A visual inspection of both AGB maps illustrates the general agreement as to biomass increases and decreases, but some highlighted areas in Figure 7 indicate apparent disparities. The mean AGB difference between the USFS and USEPA products was large (64.1 Mg/ha) at the 250 m pixel resolution (Table 11). This difference can be attributed to scaling the USEPA product from the original modeled 15 m resolution to the USFS modeled resolution of 250 m. The 15 m USEPA AGB product accounted for biomass across 52.4 percent of PR. Scaling to the 250 m resolution increased the forested component to 89.7 percent of the land area, effectively diluting the original mean biomass estimate of 72.6 Mg/ha to 42.1 Mg/ha.

Comparisons of biomass-to-biomass pixels (250 m) at the "percent forest" pixel stratification levels as defined, indicated the USEPA AGB was 51.2 Mg/ha greater on average than the USFS AGB across the five 20 percent canopy density intervals (Table 12). Low agreement (R^2 = 0.16–0.22) were observed between both AGB products across all 20 percent classes, with no notable differences between these 20 percent classes. Also, the difference between the sparsely forested pixels to the higher forested pixels was 3 to 5 times greater for the USEPA product than the USFS.

Extracting only biomass pixels within each of the five FIA forest type classes for both AGB products also resulted in low agreement between the two AGB products (R^2 = <0.01- 0.12) (Table 13). The Moist Forest class showed the best agreement (R^2 = 0.12) compared to the other four forest classes and the worst agreement was observed in the Lower Montane/Rain

Table 12. Correlation Between the Two AGB Products across 20 Percent Forest Density Intervals (Mg/Ha)

Forest (%) per 250 m	R2	RMSE	Pixels (n)	EPA AGB (μ/σ)	USFS AGB (μ/σ)	USFS-EPA
0–20	0.16	66.0	10764	25.7/22.4	85.5/28.3	59.8
20–40	0.18	64.4	9807	37.9/25.0	94.9/30.1	57.0
40–60	0.19	59.7	12058	46.6/26.1	100.5/32.2	53.8
60–80	0.20	59.5	13164	55.1/26.6	105.0/34.1	49.9
80–100	0.22	57.4	33811	73.0/25.7	119.5/37.1	46.5
0–100	0.29	60.6	79606	55.4/30.7	106.6/36.2	51.2

Table 13. Correlation Between Two AGB Products Across PR GAP Forest Classes (Mg/ha)

Forest (%) per 250 m	R2	RMSE	Pixels (n)	EPA AGB (μ/σ)	USFS AGB (μ/σ)	USFS-EPA
Mangrove	0.08	58.6	390	70.9/13.6	111.3/43.7	40.4
Dry	0.03	52.8	2069	71.3/14.7	110.5/34.7	39.3
Moist	0.12	47.4	10858	87.6/14.7	119.8/36.9	32.2
Wet/Rain	0.02	62.2	10316	87.6/15.7	141.3/29.2	53.8
L. Montane	0.01	66.2	1968	92.6/25.3	146.5/27.8	53.8

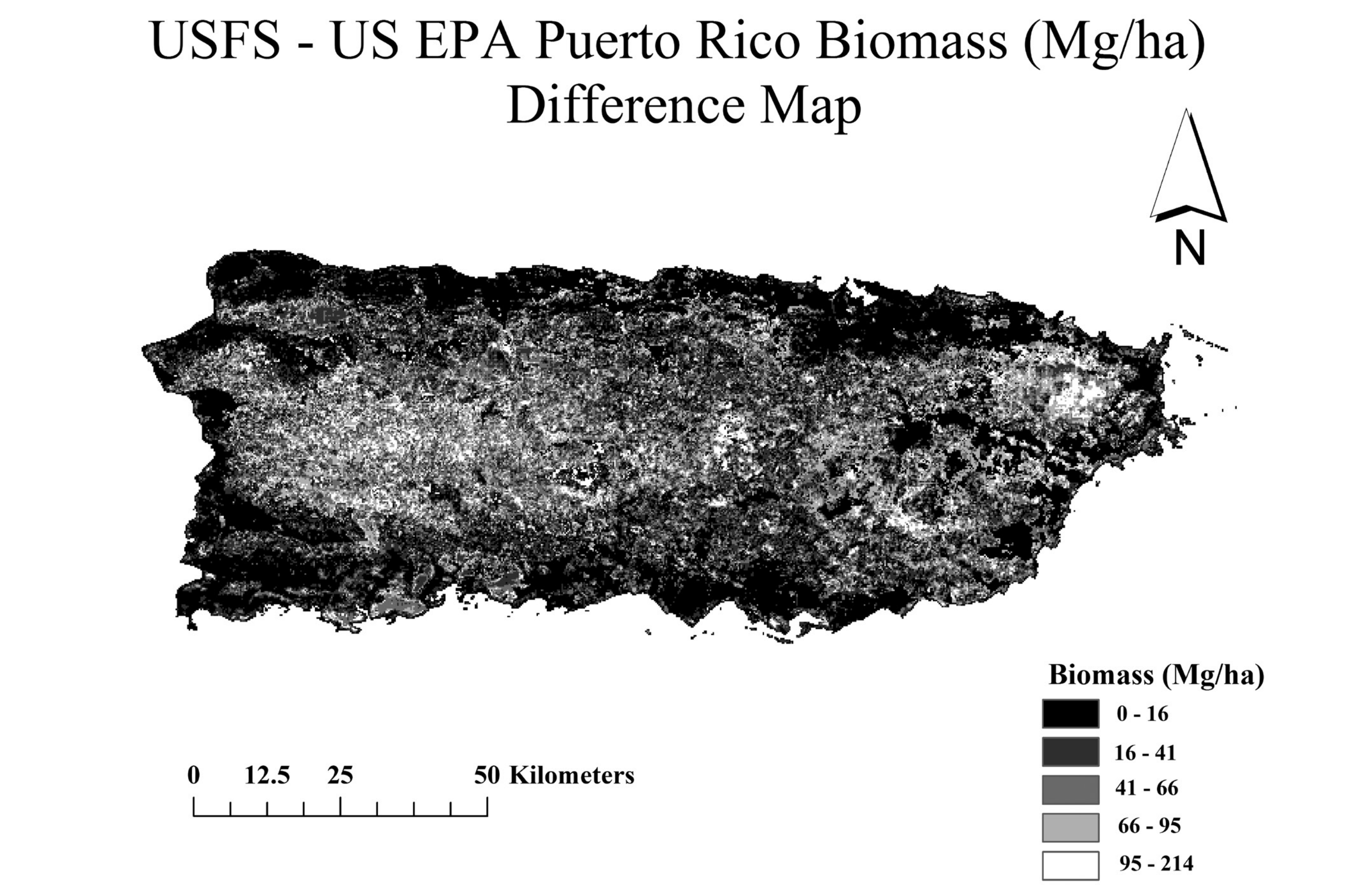


Figure 8. AGB difference map for the main island of Puerto Rico (2000).

Forest (R^2 = <0.01). This agreed with the AGB difference map, especially on the north-eastern part of the main island in the El Yunque National Forest (Figure 8). The mean difference (USFS minus USEPA) was 40.73 Mg/ha with a range of 213.7 Mg/ha. The frequency distribution showed that a majority of the differences occurred in the 0 to 40 Mg/ha range (Figure 9). Of note is that the USFS AGB Mg/ha was greater than the USEPA AGB in over 59.3 percent of the 250 m pixels.

AGB Transect Comparisons

Four 90 km west-east transects (Figure 10) applied to the USFS less USEPA AGB maps revealed greater variability in the two centrally located transects when compared to the two external (T1 and T4) transects (Figure 11). This behavior may be attributed to the external transects close association with the lower elevations and milder transitions through moisture gradients than that of the internal transects. Also, these lower elevation areas were higher in biomass which trended to be overestimated with the USEPA modeling methodology. Finally, slope differences between the two datasets revealed that the USFS AGB product had higher slope values or greater degree of change between neighboring pixels when compared

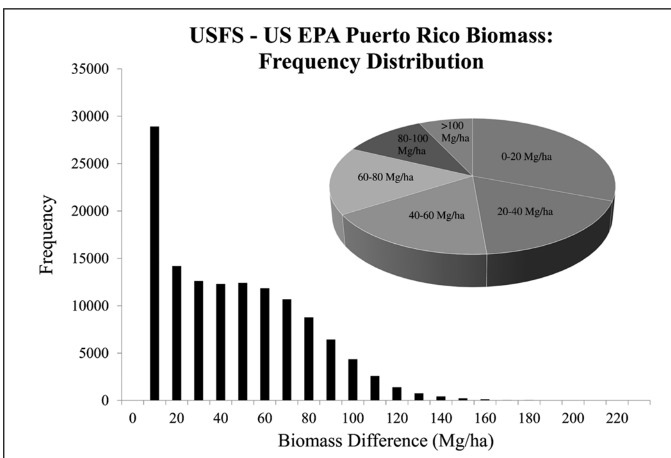


Figure 9. Frequency distribution of biomass differences across the main island of Puerto Rico.

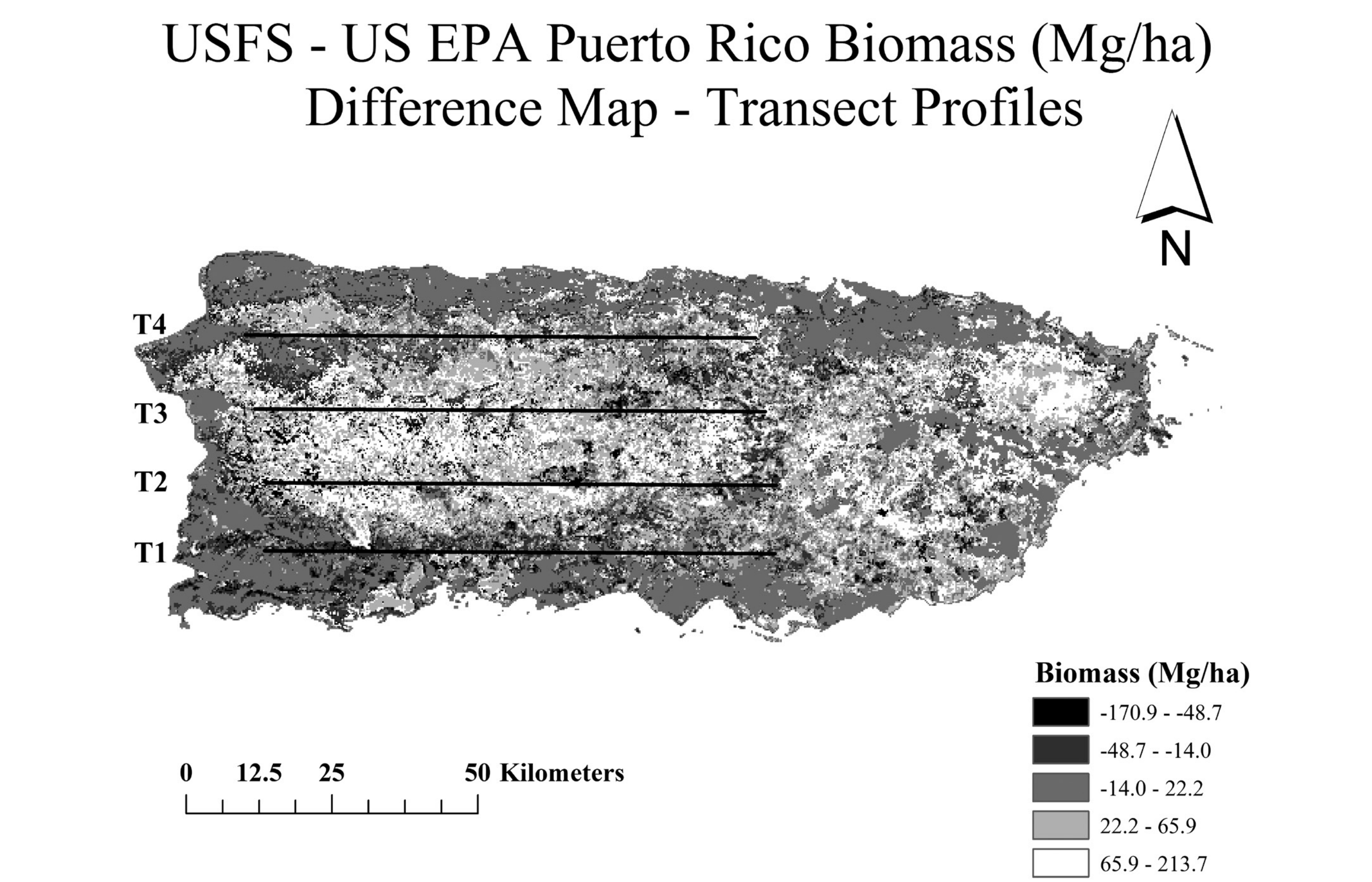


Figure 10. Transect difference profiles (positional) across Puerto Rico (Note: Positive differences indicate USFS dominant biomass; negative differences indicate USEPA dominate biomass).

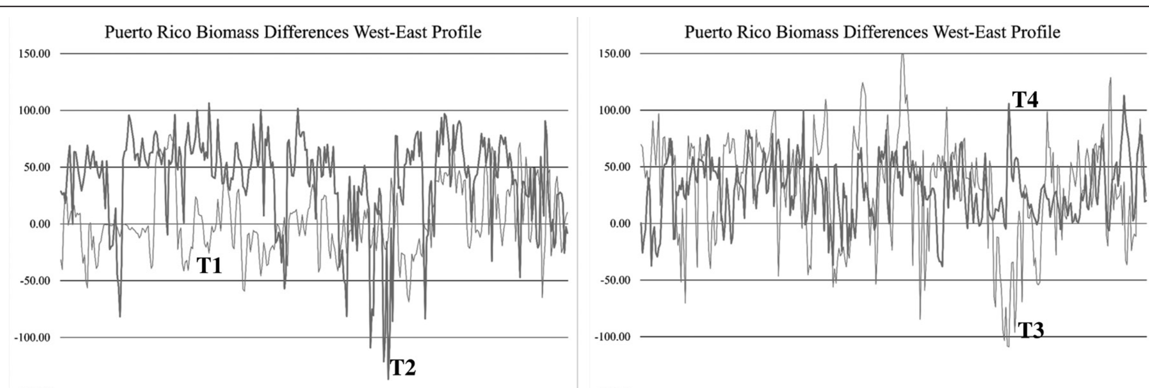


Figure 11. Transect difference profiles (graphical) across Puerto Rico (Note: Positive differences indicate USFS dominant biomass, negative differences indicate USEPA dominate biomass). (Note: Transect 1 (southern-most transect), Transect 4 (northern- most transect)

to the USEPA (Figure 12). The histogram in Figure 12 shows a positive slope distribution (mean = 3.4, SD = 6.8), where 90.0 percent of the USFS AGB pixels exceeded slope values of the USEPA product excluding slope values equal to zero. Note that 10 percent of the USFS pixel slope values exceeded 15 percent compared to the corresponding USEPA values.

Discussion

The USEPA PR AGB map was produced at 15 m spatial resolution using the RF regression tree algorithm where five primary raster-based geospatial predictor variables were used to predict one reference plot-based biomass response variable. The correlation coefficient showed good agreement ($r = 0.93$), yet the biomass model tended to underestimate high observed biomass values and overestimate low observed biomass values. Likewise, predicted MaxHt and BAWAveHt metrics showed strong agreement ($r = 0.93$, respectively) when compared to a suite of five predictor variables, yet the model also under-predicted and over-predicted canopy height in high and low observed canopy respectively, similar to Walker (2007).

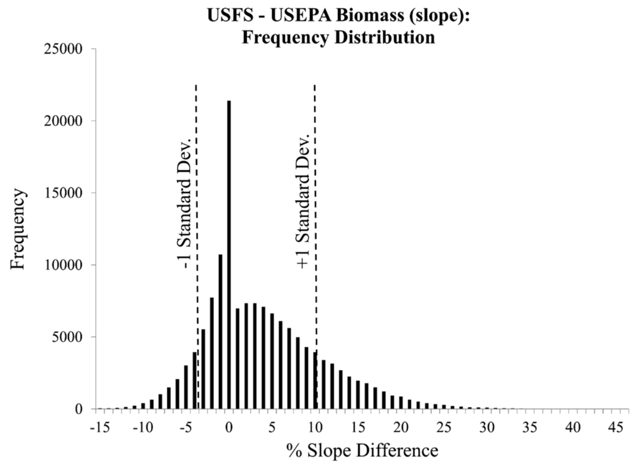


Figure 12. AGB slope distribution indicating percent change in biomass of adjacent pixels (positive values = USFS, negative values = USEPA). Note: Slope is the percent change between neighboring pixels

Internal uncertainties affecting the USEPA AGB product include canopy height anomalies as a function of DEM and SRTM error and LC biophysical properties and estimation. The SRTM-derived canopy height data was problematic producing both anomalous negative and positive values, mostly in the negative height range. The GAP cover class 'woodland/shrubland' yielded low average user and producer accuracies (52 to 73 percent), when compared to the other classes that were estimated at higher accuracies, leading to an overall 84.9 percent accuracy value. Direct comparison between the USEPA and USFS AGB products was difficult due to the initial spatial resolution differences and the definition of the threshold of how much tree cover embedded in one 250 m pixel constitutes inclusion into the biomass map. This was evident with the percent forested area increasing from 52.4 to 89.7 percent after the aggregation of the USEPA 15 m resolution to the 250 m resolution. This effectively reduced the mean forested area AGB 30.5 Mg/ha for the USEPA product. Therefore, the most appropriate comparison of overall forested AGB was between the USFS 250 m product (mean = 106.1 Mg/ha) and the USEPA 15 m product (mean = 72.6 Mg/ha), a difference of 33.5 Mg/ha. This difference was not out-of-scope when compared to other biomass comparative studies. For example, in Brazil biomass varied by more than a factor of two (Olson *et al.*, 1983; Houghton *et al.*, 2001; Gibbs, 2006)..

Given that the USFS and USEPA AGB footprints differed by approximately one order of magnitude, error estimates were affected accordingly. These spatial mismatch issues can both reduce or inflate AGB estimates, an issue addressed throughout the literature (Congalton and Plourde, 2000; Foody, 2002; Verbyla and Hammond, 1995). In the case for the USFS AGB product, the smaller FIA plot co-registered with 250 m product may not have represented the forest structure extending beyond its footprint. Under this scenario, error estimates would inflate where local forest biomass variability and/or LC heterogeneity are/is high (Blankard *et al.*, 2008). Contrary, the smaller USEPA footprint compared to the corresponding FIA footprint may not have captured variability occurring at the larger FIA plot level. It should be noted that the underlying spatial resolution of both the USFS (500 m) and USEPA (30 m) AGB maps also had data inputs larger than the base spatial scale.

Blackard *et al.* (2008) reported reasonable accuracy in the development of the forest mask where 79 percent accuracies were achieved. AGB correlation with ground referenced data was $r = 0.92$, similar to the USEPA correspondence ($r = 0.93$). The USFS found that the forest cover variable rated highest in value, especially in areas <60 percent forested. They also commented on the weakness of this dataset in the disjoint between maximums (proportional tree cover vs. forest biomass) where proportional tree cover reaches its maximum before forest biomass (Blackard *et al.*, 2008). The USEPA AGB product ranked the elevation variable as the most important within the modeling framework. Canopy density was also important and was summarily included, but it wasn't the rated the highest possibly due to canopy densities >60 percent over 51.6 percent of the island.

Conclusions

The USEPA AGB product provided internally reliable biomass estimates across the entire island of PR. We developed our product at 15 m to support the development of landscape watershed predictor metrics for sediment and nutrient loadings associated with stream reaches. Given differences in initial spatial resolution, modeling classifiers and input variables used, and forest/non-forest delineation differences, the disjoint between the two products was evident. This is also proved by the very low correlation values ($R^2 = 0.01$ to 0.12) and the high mean difference values (40.4 to 53.8 Mg/ha) between both products across the five GAP LC classes. However, the differences were within the range seen in other global and regional biomass studies. The significant difference in average AGB estimates between the USEPA and USFS products indicates that additional research is needed to validate the currently available AGB products for Puerto Rico.

Acknowledgments

The authors would like to thank Sam Lambert with the USFS Southern Research Station FIA Research Work Unit (Knoxville, Tennessee) for his assistance with this project. Also, special thanks to Eileen Helmer (USFS) in providing information on the USFS AGB product methods. Alessandro Baccini from Woods Hole, Massachusetts also provided further detail into the NBCD2000 methodology. Finally, the authors would like to thank the four anonymous reviewers who gave pertinent feedback to this manuscript. The USEPA funded and conducted the research described in this paper. Although this work was reviewed by EPA and has been approved for publication, it may not necessarily reflect official Agency policy. Mention of any trade names or commercial products does not constitute endorsement or recommendation for use.

References

Abuyuan, A., I. Hawken, M. Newkirk and R. Williams, 1999. Waste equals food: Developing a sustainable agriculture support cluster for a proposed resource recovery park in Puerto Rico, *Yale School of Forestry and Environmental Studies Bulletin*, 106:303–349.

Asner, G.P., R.F. Hughes, J. Mascaro, A.L. Uowolo, D.E. Knapp, J. Jacobson, T. Kennedy- Bowdoin, and J.K. Clark, 2011. High-resolution carbon mapping on the million-hectare Island of Hawaii, *Frontiers in Ecology and the Environment*, 9(8):434–439.

Baccini, A., and G.P. Asner, 2013. Improving pantropical forest carbon maps with airborne LiDAR sampling, *Carbon Management*, 4(6):591–600.

Bamler, R., 1999. The SRTM Mission: A world-wide 30 m resolution DEM from SAR Interferometry in 11 days, *Photogrammetric week '99* (D. Fritch and R. Spiller Wichmann, editors), Heidelberg (1999), pp. 145–154.

Basuki, T.M., P.E. van Laake, A.K. Skidmore, and Y.A. Hussin, 2009. Allometric equations for estimating above-ground biomass in tropical lowland *Dipterocarp* forests, *Forest Ecology and Management*, 257:1684–1694.

Bayley, S.E., and J.K. Guimond, 2008. Effects of river connectivity on marsh vegetation community structure and species richness in montane floodplain wetlands in Jasper National Park, Alberta, Canada, *Ecoscience*, 15:377–388.

Blackard, J., M. Finco, E. Helmer, G. Holden, M. Hoppus, D. Jacobs, A. Lister, G. Moisen, M. Nelson, R. Riemann, B. Ruefenacht, D. Salajanu, D. Weyermann, K. Winterberger, T.

Brandeis, R. Czaplewski, R. McRoberts, P. Patterson, and R. Tymcio, 2008. Mapping U.S. forest biomass using nationwide forest inventory data and MODIS-based information, *Remote Sensing of Environment*, 112(4):1658–1677.

Brandeis, T.J., 2003. Puerto Rico's forest inventory: Adapting the forest inventory and analysis program to a Caribbean island, *Journal of Forestry*, 101:8–13.

Brandeis, T., 2006. Forest Inventory and Analysis Factsheet Puerto Rico 2003, URL: *http://srsfia2.fs.fed.us/states/pr/PR_Factsheet.pdf* (last date accessed: 07 February 2017).

Brandeis, T.J., M.B. Delaney, R. Parresol, and L. Royer, 2006. Development of equations for predicting Puerto Rican subtropical dry forest biomass and volume, *Forest Ecology and Management*, 233:133–142.

Brandeis, T.J., E.H. Helmer and S.N. Oswalt, 2007. The status of Puerto Rico's forests, 2003, *Resource Bulletin SRS-119*, Asheville, North Carolina, U.S. Department of Agriculture Forest Service, Southern Research Station.

Breiman, L., 2001. Random forests, *Machine Learning*, 45(1):5–32.

Brown, S., A.J.R. Gillespie, and A.E. Lugo, 1989. Biomass estimation methods for tropical forests with applications for forest inventory data, *Forest Science*, 35:881–902.

Cartus, O., J. Kellndorfer, W. Walker, C. Franco, J. Bishop, L. Santos, J. María and M. Fuentes, 2014. A National, Detailed Map of Forest Aboveground Carbon Stocks in Mexico, *Remote Sensing*, 6:5559–5588.

Cintrón, G., and Y. Schaeffer-Novelli, 1984. Caracteristicas y desarrollo estructural de los manglares de Norte y Sur América, *Ciencia Interamericana*, 25:4–15.

Congalton, R.G., and L.C. Plourde, 2000. Sampling methodology, sample placement, and other important factors in assessing the accuracy of remotely sensed forest maps, *Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Delft University, Amsterdam, pp. 117–124.

Curkendall, D., E. Fielding, T. Cheng, and J. Pohl, 2003. A computational grid based system for continental drainage network extraction using SRTM digital elevation models, *IEEE ICPP Workshops*, pp. 181– 190.

Dubayah, R.O., and J.B. Drake, 2000. Lidar remote sensing for forestry, *Journal of Forestry*, 98(6):44–46.

Foody, G.M., 2002. Status of land cover classification accuracy assessment, *Remote Sensing of Environment*, 90:185–201.

Gesch, D.B., M.J. Oimoen, and G.A. Evans, 2014. Accuracy assessment of the U.S. Geological Survey National Elevation Dataset, and comparison with other large-area elevation datasets - SRTM and ASTER: U.S. Geological Survey Open-File Report 2014–1008, 10 p., URL: *https://pubs.usgs.gov/of/2014/1008/pdf/ofr2014-1008.pdf*, (last date accessed: 07 February 2017).

Gibbs, H.K. 2006. Olson's major world ecosystem complexes ranked by carbon in live vegetation: An updated database using the GLC2000 land cover product, NDP-017b. ORNL-CDIAC, URL: *http://cdiac.ornl.gov/epubs/ndp/ndp017/ndp017b.html* (last accessed 07 February 2017).

Gould, W., C. Alarcón, B. Fevold, M.E. Jiménez, S. Martinuzzi, G. Potts, M. Solórzano, and E. Ventosa, 2008. *Puerto Rico Gap Analysis Project - Final Report*, USGS, Moscow, Idaho and the USDA Forest Service International Institute of Tropical Forestry, Río Piedras, PR.

Harris, N.L., S. Brown, S.C. Hagen, S.S. Saatchi, S. Petrova, W. Salas, M.C. Hansen, P.V. Potapov, and A. Lotsch, 2012. Baseline map of carbon emissions from deforestation in tropical regions, *Science*, 336:1573–1576.

Helmer, E.H., O.M. Ramos-González, T. del M. López, and W. Díaz, 2002. Mapping the forest type and land cover of Puerto Rico - A component of the Caribbean biodiversity hotspot, *Caribbean Journal of Science*. 38(3-4):165–183.

Hoffman, J., and D. Walter, 2006. How complementary are SRTM-X and -C band digital elevation models?, *Photogrammetric Engineering & Remote Sensing*, 72(3):261–268.

Houghton, R.A., K.T. Lawrence, J.L. Hackler, and S. Brown, 2001. The spatial distribution of forest biomass in the Brazilian Amazon: A comparison of estimates, *Global Change Biology*, 7(7):731–46.

Huang, C., L. Yang, B. Wylie, and C. Homer, 2001. A strategy for estimating tree canopy density using Landsat 7 ETM+ and high resolution images over large areas, *Proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry*, 05-07 November, Denver, Colorado, unpaginated CD-ROM.

Iiames, J.S., R.G. Congalton, T.E. Lewis, and A.E. Pilant, 2015. Uncertainty analysis in the creation of a fine-resolution leaf area index (LAI) reference map for validation of moderate resolution LAI products, *Remote Sensing*, 7:1397–1421.

Imhoff, M.L., 1995. Radar backscatter and biomass saturation: Ramifications for global biomass inventory, *IEEE Transactions on Geoscience and Remote Sensing*, 33:(2):511–518.

Kellndorfer, J., W. Walker, L. Pierce, C. Dobson, J.A. Fites, C. Hunsaker, J. Vona, and M. Clutter, 2004. Vegetation height estimation from shuttle radar topography mission and national elevation datasets, *Remote Sensing of Environment*, 93(3):339–358.

Kellndorfer, J., W. Walker, E. LaPoint, M. Hoppus, and J. Westfall, 2006. Modeling height, biomass, and carbon in US forests from FIA, SRTM, and ancillary national scale data sets, *Proceedings of the 2006 IEEE International Symposium on Geoscience and Remote Sensing*, 31 July - 04 August, Denver, Colorado., pp. 3591–3594.

Koch, B., 2010. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment, *ISPRS Journal of Photogrammetry and Remote Sensing*, 65:581–590.

Larsen, M.C., and A. Simon, 1993. A rainfall intensity-duration threshold for landslides in a humid-tropical environment, Puerto Rico, *Geografiska Annaler*, Series A. *Physical Geography*, 75:13–23.

Lefsky, M.A., W.B. Cohen, D.J. Harding, G.G. Parker, S.A. Acker, and S.T. Gower, 2002. Lidar remote sensing of above-ground biomass in three biomes, *Global Ecology and Biogeography*, 11(5):393–399.

Lunetta, R.S., R.G. Congalton, L.K. Fenstermaker, J.R. Jensen, K.C. McGuire, and L.R. Tinney, 1991. Remote sensing and geographic information systems data Integration: Error sources and research issues, *Photogrammetric Engineering & Remote Sensing*, 57(6):677–687.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik, 2013. Cluster analysis basics and extensions, *R Package version 1.14.4.*

Marks, W.L., J.S. Iiames, R.S. Lunetta, S. Khorram, and T.H. Mace, 2014. Basal area and biomass estimates of loblolly pine stands using L-band UAVSAR, *Photogrammetric Engineering & Remote Sensing*, 80(1):33–42.

Mitchell, M.J., G. Lovett, S. Bailey, F. Beall, D. Burns, D. Buso, T.A. Claire, F. Courchesne, L. Duchesne, C. Eimers, and I. Fernandez, 2011. Comparisons of watershed sulfur budgets in southeast Canada and northeast US: New approaches and implications, *Biogeochemistry*, 103(1-3):181–207.

Murphy, S.F., R.F. Stallard, M.C. Larsen, and W.A. Gould, Land cover of four watersheds in eastern Puerto Rico, *Water Quality and Landscape Processes of Four Watersheds in Eastern Puerto Rico*, USGS Professional Paper 1789, URL: *https://pubs.usgs.gov/pp/1789/PP1789.pdf* (last date accessed: 07 February 2017).

Nowak, D.J., and E.J. Greenfield, 2010. Evaluating The National Land Cover Database tree canopy and impervious cover estimates across the conterminous United States: A comparison with photo-interpreted estimates, *Environmental Management*, 46:378–390.

Olson, J.S., J.A. Watts, and L.J. Allison, 1983. *Carbon in Live Vegetation of Major World Ecosystems*, ORNL-5862, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 164 p.

Pan Y., R.A Birdsey, J. Fang, R. Houghton, P.E. Kauppi, W.A. Kurz, O.L. Phillips, A. Shvidenko, S.L. Lewis, J.G. Canadell, P. Ciais, R.B. Jackson, S.W. Pacala, A.D.

McGuire, S. Piao, A. Rautiainen, S. Sitch, and D. Hayes, 2011. A large and persistent carbon sink in the world's forests, *Science*, 333(6045):988–993.

Quinlan, J.R., 1986. Induction of decision trees, *Machine learning*, 1(1):81–106.

Quinlan, J.R., 1994. C4.5: Programs for Machine Learning, *Morgan Kaufman Publishers,* San Mateo, California.

Rosen, P., M. Eineder, B. Rabus, E. Gurrola, S. Hensley, W. Knoepfle, H. Breit, A. Roth, and M. Werner, 2001a. SRTM mission - Cross comparison of X and C band data properties, *Proceedings of IGARSS*, July, Sydney, Australia, 2:751–753.

Rosen, P., S. Hensley, E. Gurrola, F. Rogez, S. Chan, S., J. Martin, and E. Rodriguez 2001b, SRTM C-band topographic data: Quality assessments and calibration activities, *Proceedings of IGARSS*, July, Sydney, Australia, 2:739–741.

Sarabandi, K., and Y.C. Lin, 2000. Simulation of interferometric SAR response for characterizing the scattering phase center statistics of forest canopies, *IEEE Transactions on Geoscience and Remote Sensing*, 39:115–125.

Scatena, F.N., W.L. Silver, T. Siccama, A. Johnson, and M.J. Sanchez, 1993. Biomass and nutrient content of the Bisley Experimental Watershed, Luquillo Experimental Forest, Puerto Rico, before and after Hurricane Hugo, 1989, *Biotropica*, 25:15–27.

Simard, M., K. Zhang, V.H. Rivera-Monroy, M.S. Ross, P.L. Ruiz, E. Castañeda-Moya, R.R.

Twilley, and E. Rodriguez, 2006. Mapping height and biomass of mangrove forests in Everglades National Park with SRTM elevation data, *Photogrammetric Engineering & Remote Sensing*, 72(3):299–311.

Sexton, J.O., T. Bax, P. Siqueira, J.J. Swenson, and S. Hensley, 2009. A comparison of lidar, radar, and field measurements of canopy height in pine and hardwood forests of southeastern North America, *Forest Ecology and Management*, 257(3):1136–1147.

Smith, W.B., 2002. Forest inventory and analysis: A national inventory and monitoring program, *Environmental Pollution*, 116:S233–S242.

Smith, B., and D. Sandwell, 2003. Accuracy and resolution of Shuttle Radar Topography Mission data, *Geophysical Research Letters*, 30:1467–1470.

Treuhaft, R.N., B.D. Chapman, J.R. dos Santos, F.G. Goncalves, L.V. Dutra, P.M.L.A. Graca and J.B. Drake, 2009. Vegetation profiles in tropical forests from multibaseline interferometric synthetic aperture radar, field, and lidar measurements, *Journal of Geophysical Research*, 114:1–16.

Verbyla, D.L., and T.O. Hammond, 1995. Conservative bias in classification accuracy assessment due to pixel-by-pixel comparison of classified images with reference grids, *Remote Sensing*, 16(3):581–587.

Walker, W.S., J.M. Kellndorfer, E. LaPoint, M. Hoppus, and J. Westfall, 2007. An empirical InSAR-optical fusion approach to mapping vegetation canopy height, *Remote Sensing of Environment*, 109:482–499.

Warne, A.G., R.M.T. Webb, and M.C. Larsen, 2005. *Water, Sediment, and Nutrient Discharge Characteristics of Rivers in Puerto Rico, and Their Potential Influence on Coral Reefs*, U.S. Geological Survey Scientific Investigations Report 2005-5206, Reston, Virginia.

Weaver, P.L., and A.J. Gillespie, 1992. Tree biomass equations for the forests of the Luquillo Mountains, Puerto Rico, *The Commonwealth Forestry Review*, 71:35–39.

Zolkos, S.G., S.J. Goetz, and R. Dubayah, 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing, *Remote Sensing of Environment*, 128:289–298.

(Received 16 January 2016; accepted 15 August 2016; final version 18 January 2017)