Remote Sensing of Environment 121 (2012) 426–442

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

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

Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa)

João M.B. Carreiras [a,*], Maria J. Vasconcelos [a], Richard M. Lucas [b]

[a] Tropical Research Institute (IICT), Travessa do Conde da Ribeira, 9, 1300-142 Lisbon, Portugal
[b] Institute of Geography and Earth Sciences, Aberystwyth University, Aberystwyth, Ceredigion, Wales, SY23 3DB, Wales, UK

ARTICLE INFO

Article history:
Received 4 August 2011
Received in revised form 8 February 2012
Accepted 11 February 2012
Available online 22 March 2012

Keywords:
Guinea-Bissau
West Africa
Aboveground biomass
REDD+
Remote sensing
ALOS PALSAR
Bagging stochastic gradient boosting (BagSGB)

ABSTRACT

Guinea-Bissau is one of the poorest countries in the world with a large proportion of its population living in rural areas. While industry is limited, over 70% of the territory is covered by forests, which can potentially be used to attract investment through forest-based projects that promote reductions in carbon emissions and sustainable management. These can be leveraged by producing accurate maps of forest aboveground biomass (AGB) at national level and by developing cost-effective mapping methods that allow reliable future updating for management and engagement in international mechanisms such as the United Nations (UN) Collaborative Program on Reducing Emissions from Deforestation and Forest Degradation (REDD) in developing countries.

Using data from Japan's Advanced Land Observing Satellite (ALOS) Phased Array L-band Synthetic Aperture Radar (PALSAR), this study compared a semi-empirical and machine learning algorithm, with the latter based on bagging stochastic gradient boosting (BagSGB), for retrieving the AGB of woody vegetation thereby supporting estimation of national carbon stocks. AGB was estimated by using measurements of tree size collected from112 forest plots during two field campaigns (2007 and 2008) as input to published allometric equations.

The BagSGB outperformed the semi-empirical algorithm, resulting in a coefficient of correlation (R) between observed and cross-validation predicted forest AGB values of 0.95 and in a root mean square error (RMSE) of 26.62 Mg ha^{-1}. Furthermore, the BagSGB model produced also a measure of forest AGB prediction variability (coefficient of variation) on a pixel-by-pixel basis, with values ranging from 7 to 250% (mean = 42%). An estimate of total forest AGB carbon stock of 96.93 Mt C was obtained in this study for Guinea-Bissau, with a mean forest AGB value of 65.17 Mg ha^{-1}. Although the mean error associated with this forest AGB map is still undesirably high, several issues were addressed. The heterogeneity of forest structural types, presence of palm trees, and dimension and type of field plots were identified as potential source of uncertainty that must be tackled in future studies. This study represents a step forward regarding the information currently available for Guinea-Bissau.

1. Introduction

Mapping and understanding the spatial distribution of forest aboveground biomass (AGB) is an important and challenging task (e.g., Baccini et al., 2008; Lucas et al., 2010; Saatchi et al., 2011a). As it relates to the carbon stocks of a given ecosystem, these maps can be used to monitor forests and capture national deforestation processes, forest degradation, and the effects of conservation actions, sustainable management and enhancement of carbon stocks. The importance of accurately reporting the carbon content (biomass) of forested lands has been acknowledged and emphasized by several studies (e.g., Canadell et al., 2004; Falkowski et al., 2000; Heimann and Reichstein, 2008; Houghton, 2005; Le Quere et al., 2009; Loarie et al., 2009; Schimel et al., 2001; Schulze, 2006) and is a requirement of international conventions (e.g., the United Nations (UN) Framework Convention on Climate Change (UNFCCC); UN, 1992). Such information is critical as it forms the basis of reporting mechanisms developed under the UNFCCC post-Kyoto Protocol (UN, 1998) and particularly the initiative focusing on Reducing Emissions from Deforestation and forest Degradation (REDD; http://www.unfccc.redd.org) in developing countries (FAO et al., 2008) and REDD+ which additionally considers mitigation and conservation actions, sustainable management of forests and enhancement of carbon stocks (Campbell, 2009; UNFCCC, 2011).

Numerous studies have demonstrated that a relationship exists between the AGB of forests and low frequency (L- and P-band) Synthetic Aperture Radar (SAR) data. The majority have focused on closed forests in boreal, temperate and tropical regions (e.g., Austin

* Corresponding author. Tel.: +351 213616340x214.
E-mail address: jmbcarreiras@iict.pt (J.M.B. Carreiras).

0034-4257/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2012.02.012

et al., 2003; Dobson et al., 1992, 1995; Fransson and Israelsson, 1999; Hoekman and Quinones, 2000; Imhoff, 1995; Le Toan et al., 1992; Lucas et al., 2010; Luckman et al., 1997; Saatchi et al., 2007; Santos et al., 2002) and the majority have reported saturation at approximately 60–100 Mg ha^{-1} and 100–150 Mg ha^{-1} for L- and P-band respectively. However, robust retrieval remains a significant challenge, particularly for higher biomass forests (i.e., approaching and above the saturation level) and because of variations in structure and environmental conditions. For example, Lucas et al. (2010) highlighted the sensitivity of ALOS PALSAR HH polarization, and to a lesser extent HV polarization, data to surface moisture conditions, recommending that AGB is best retrieved using data acquired during periods of minimal soil/surface moisture and rainfall. Variability in the relationships was also attributed to differences in vegetation structure (e.g., size and density) as a function of forest type (e.g., sparse and open woodlands, closed forest) and regrowth stage (remnant versus secondary). In a preceding study (Lucas et al., 2006), the dynamic range over which AGB could be estimated was shown to vary as a function of incidence angle [based on National Aeronautics and Space Administration (NASA) Jet Propulsion Laboratory (JPL) Airborne SAR (AIRSAR)] and the acquisition of data at HV polarization at larger (>45°) incidence angles advocated. As one solution, Saatchi et al. (2011b) suggested that scatter in the relationship could be reduced by extracting backscatter data from larger areas around known field plot locations. Through these and other studies, a better understanding of the reasons for variability and scatter in the AGB–SAR relationships is being obtained, which is ultimately leading to better retrieval.

In most cases the retrieval of forest AGB from lower frequency SAR data has been achieved though inversion of relationships with single channel (typically HV) data. However, this study sought to establish whether a semi-empirical model, based on that of Saatchi et al. (2007) and a machine learning algorithm (based on bagging stochastic gradient boosting, BagSGB) could be used as alternatives to estimation of AGB. The methods were evaluated using a mosaic of 26 ALOS PALSAR fine beam dual (FBD) (HH and HV polarizations, 34.3° off-nadir angle) scenes acquired in 2008 and covering the entire territory of Guinea-Bissau (West Africa), and forest AGB estimates resulting from tree size measurements taken from 112 plots in 2007 and 2008, and distributed across a range of forest types (i.e., closed forest, open forest, savanna woodland and mangrove).

The paper is structured as follows. Section 2 provides the framework and the scientific background of the research, including a brief description of the study area, field data collection, and estimation of biomass at the plot level. Section 3 describes the remote sensing data used in this study, as well as the methods used to process the ALOS PALSAR FBD data and to establish relationships with AGB. The results and corresponding discussion are presented in Section 4 and the study is concluded in Section 5.

2. Framework and scientific background

2.1. Context

Reducing tropical deforestation has driven the international community to the establishment of investment mechanisms and market-based carbon transactions targeted to avoiding deforestation while promoting development. These mechanisms are highly relevant for developing tropical countries (e.g., REDD, REDD+; FAO et al., 2008; UNFCCC, 2011).

In West and Central Africa, the mean annual rate of forest loss was 0.46% between 2005 and 2010 (FAO, 2010). In an effort to reduce this rate through mechanisms such as REDD and REDD+, countries in the region need to have access to the carbon market which can be achieved in part through the production of reliable information on carbon stocks in vegetation and changes associated with human activities (e.g., forest clearance for agriculture). By establishing baselines of carbon stocks and land cover, carbon emissions from land use and land cover change can be quantified and impacts of different management activities can be compared. Monitoring of changes in carbon stocks can also be undertaken to demonstrate the relative impacts of land management activities and maps generated can support accreditation. Methods that allow routine mapping of AGB using remote sensing data are also desirable, particularly given the high costs associated with field survey and the requirement of investors in forest-based products for Measurement, Reporting and Verification (MRV).

Guinea-Bissau, a country located in West Africa between 10°–13° N and 13°–17° W (Fig. 1), has an area of approximately 35 000 km^2 and a total population (by 2008) of 1 575 000 habitants (FAO, 2010). It is among the ten poorest countries in the world, with a per capita gross domestic product of US$537 and 70% of its population depending directly on natural resources for subsistence (FAO, 2010). Since there is almost no industry, sustainable land use/land cover management is the only means by which the country can immediately access the carbon market and obtain investment benefits for clean development. Having ratified the Kyoto Protocol in 2005, Guinea-Bissau is therefore qualified to participate in the Clean Development Mechanisms (CDM) (Intergovernmental Panel on Climate Change, IPCC, 2007) and must report carbon emissions annually to the UNFCCC. To assist Guinea-Bissau with this endeavor, the CarboVeg-GB project (http://carboveg-gb.dpp.pt) sought to map historical deforestation and land cover change during the period 1990–2007 and map carbon stocks associated with forest vegetation types, thereby facilitating estimation of carbon emissions associated with subsequent degradation.

2.2. Field measurements

As part of the CarboVeg-GB project, field data collection was based on a stratified sampling methodology, using a land cover map obtained from a supervised classification of Landsat 5 Thematic Mapper (TM) data of 2007 (Oom et al., 2009). From the map, the area of each forest class (i.e., closed forest, open forest, savanna woodland and mangrove) was determined and the number of field plots collected in each in 2007 calculated. Additional forest plots were then measured in 2008 to increase, where necessary, the number needed for each forest class in order to achieve a desired precision level (in this case, an error level of 10% expressed as the 95% confidence interval of the mean).

The location of each plot was randomly selected based on a 250×250 m grid overlaid over the entire territory, with a random origin. Two field campaigns took place between April 4 and April 21 (2007), and March 9 and April 6 (2008), with the distribution of plots as a function of forest type and location shown in Table 1 and Fig. 1. All measurements followed a three-nest sampling plot methodology (4, 14, and 20 m concentric sub-plots), which was designed to optimize and accommodate measurements of stands (forest and palm trees) with a wide range of tree diameters (Pearson et al., 2007). In each sample plot, individual live forest trees were measured for diameter (at breast height, 1.30 m, DBH) and height, and each forest tree was identified to species (Table 2). However, forest trees were measured according to the sub-plot nest sizes: 4 m (trees with DBH>5 cm), 14 m (trees with DBH>20 cm) and 20 m (trees with DBH>50 cm). All palm trees were measured, but only the stem height and species identification were recorded. Where dead trees were encountered, the diameters at the base and top (where accessible) and stem height were recorded. Examples of forest types measured are illustrated in Fig. 2.

2.3. Estimation of AGB

At the individual tree-level, the AGB was estimated using the allometric equations of Chave et al. (2005), which were specific to

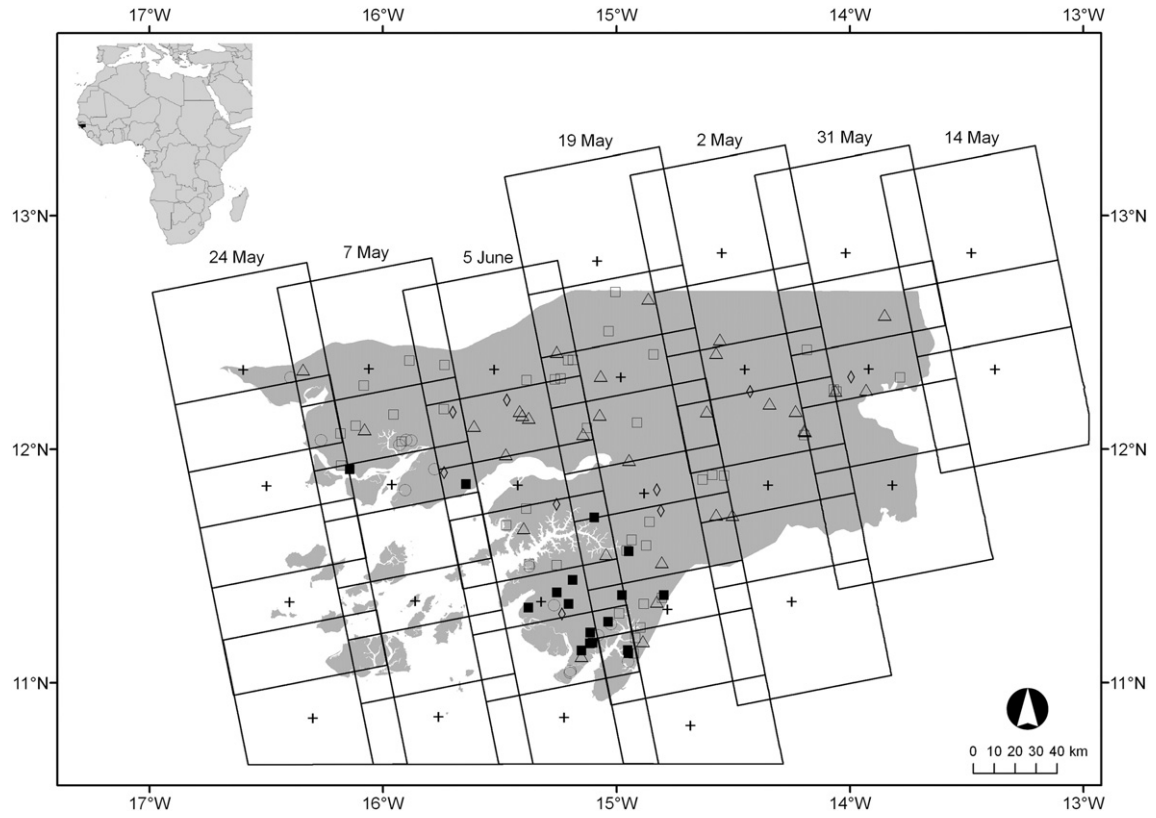

Fig. 1. Location of Guinea-Bissau in Africa (upper-left inset), and a zoom over the country showing the location of each ALOS PALSAR FBD scene (with scene center, +), including the acquisition date of each ascending pass. Also, the location of each measured field plot is shown (■—closed forest, □—open forest, Δ—savanna woodland, ○—mangrove, and ◊—non-forest).

tropical species, and tree size measurements as inputs. Separate equations as a function of climate and primarily the mean monthly evapotranspiration (ET) and rainfall (R) were developed by Chave et al. (2005), with these being wet (ET > R less than one month per year), moist (ET > R more than one month and less than five months per year), and dry (ET > R more than five months per year) forests. These criteria basically defined the extent of the growing period (when R is greater than ET) as a proportion of the year and on a monthly basis. Thus, the same criteria can be rearranged in terms of growing period (GP) as (i) wet (GP > 11/12), (ii) moist (11/12 > GP > 7/12), and (iii) dry (GP < 7/12).

2.3.1. Estimation of GP

To estimate the duration of the GP for Guinea-Bissau, which was required to determine the most appropriate allometrics to use, information on the land surface state and flux parameters generated by the Global Land Data Assimilation System (GLDAS) [Goddard Earth Sciences (GES) Data and Information Services Center (DISC), NASA] were considered (Rodell et al., 2004). GLDAS uses multiple land surface models (uncoupled with the atmosphere), integrates a high number of observation based data, and executes globally at spatial resolutions of 0.25° and 1° (Kumar et al., 2006). Currently, GLDAS uses four land surface models (LSMs): Mosaic, Noah, Community Land Model (CLM), and Variable Infiltration Capacity (VIC). The 1° spatial resolution dataset was used as this had a higher temporal coverage (1979–2008) compared to the 0.25° dataset (2000–present). Monthly rather than 3-hourly resolution data were used, with the former generated from the latter, because of interest in estimating the growth season on a monthly basis. The parameters of interest were mean surface temperature (K), total evapotranspiration (kg m^{-2} s^{-1}),

Table 1
Distribution of plots per forest type (in parentheses the number of plots discarded due to *a posteriori* screening of each field datasheet). The minimum (Min), maximum (Max), mean, and standard deviation (Std) of the AGB values for all plots following data screening (n = 112) are also shown.

Forest type	Year		Total	AGB (Mg ha^{-1})			
	2007	2008		Min	Max	Mean	Std
Closed forest	9 (2)	15 (5)	24 (7)	46.1	355.6	143.6	79.6
Open forest	15 (5)	40 (9)	55 (14)	14.8	306.2	86.8	62.2
Savanna woodland	17 (4)	19 (1)	36 (5)	0.0	126.6	38.3	35.5
Mangrove	6	10 (2)	16 (2)	4.0	91.1	34.7	30.5
Non-forest	7	2	9	0.0	58.5	17.4	18.5
Total	54 (11)	86 (17)	140 (28)	0.0	355.6	69.9	66.2

Table 2
Measurements made during field work. Other collected data included counting small trees (i.e., 0 < DBH < 5), downed dead wood measurements (in two 40 meter transects), collecting log wood sample with a gimlet (for wood specific density estimation), collecting and weighting litter sampling, and collecting and weighting dead wood samples.

Stand-level	Plot level	Tree-level
Geographic coordinates ([a]GPS)	Forest type classification	Species identification
Physiographic location	Tree crown cover	Tree vitality classification
Aspect	Vegetation vertical structure classification	Diameter at breast height (DBH)
Slope	Vegetation horizontal structure classification	Total height
	Litter depth	Top and basal diameter of dead trees
	Forest degradation factors (e.g., soil erosion, illegal logging, burning practices)	Stem height of palm trees

[a] Global Positioning System.

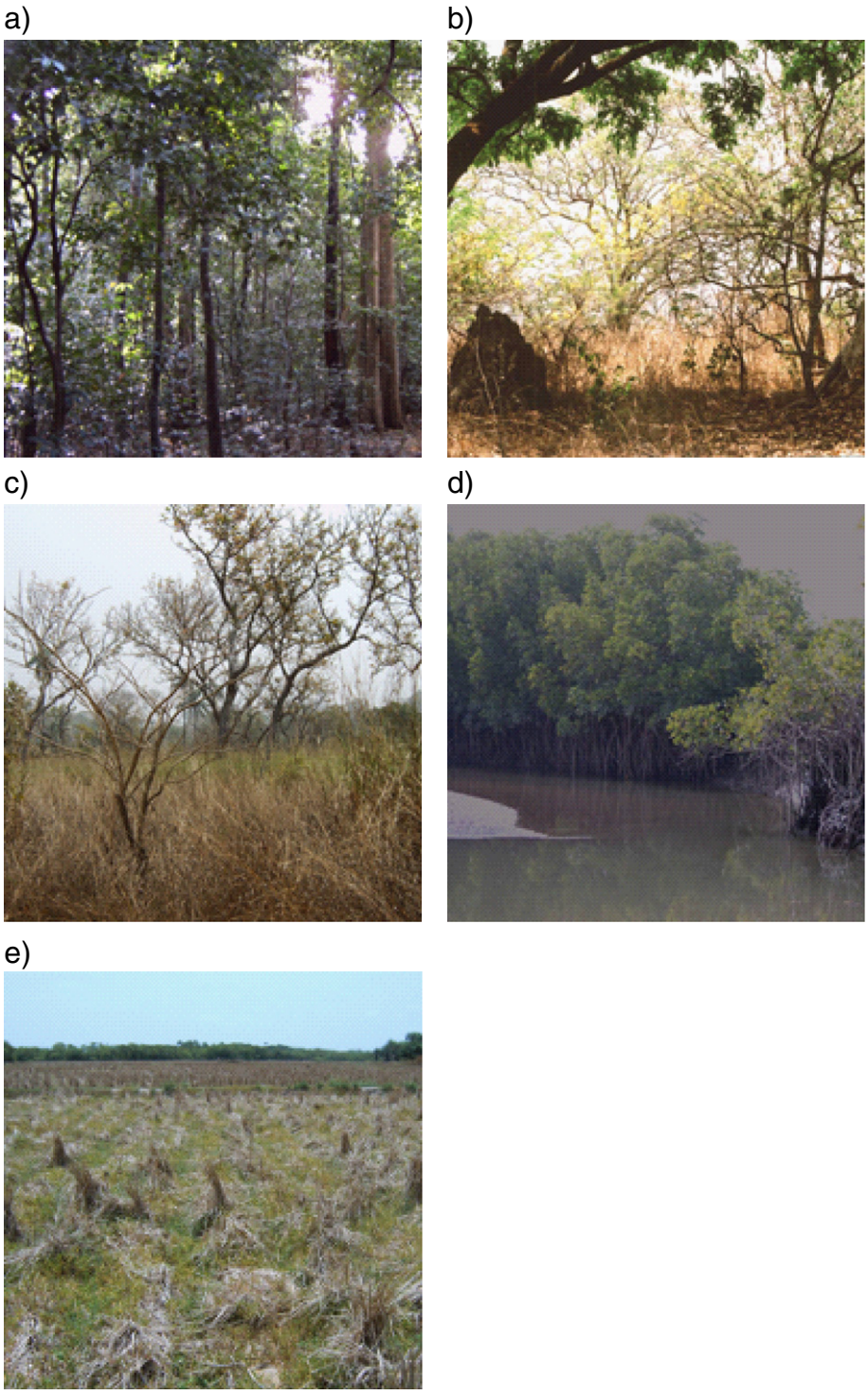

Fig. 2. Photos displaying Guinea-Bissau main forest types: a) closed forest, b) open forest, c) savanna woodland, d) mangrove, and e) non-forest.

rainfall rate ($\mathrm{kg\,m^{-2}\,s^{-1}}$), and snowfall rate ($\mathrm{kg\,m^{-2}\,s^{-1}}$). The VIC model outputs did not include mean surface temperature because the model was run in "water balance mode"; in this mode, the model does not solve the full energy balance, and not all variables are recorded. In Guinea-Bissau all precipitation was recorded as rainfall. Rates of precipitation and evapotranspiration were converted to monthly values using Eq. (1):

$$X_i = 2,592,000\, Xr_i \quad (1)$$

where X_i represents the value of variable X in the ith month, and Xr_i is the variable X rate (i.e., $\mathrm{kg\,m^{-2}\,s^{-1}}$) in the ith month. The constant is simply the product of the number of seconds in a 3-hour period (10 800), 3-hour periods in one day (8), and days in a month (30). This transformation was only applied to precipitation and evapotranspiration. Based on these data, all 12 pixels covering Guinea-Bissau had a growing period below 7/12. Consequently, all the forests of Guinea-Bissau could be considered as belonging to the dry category.

2.3.2. Allometric equations used to estimate AGB

2.3.2.1. Live forest trees. Given that forests in Guinea-Bissau can be considered as dry, the predictive model in Eq. (2) (Chave et al., 2005) was used to estimate the AGB of each live forest tree, with the exception of mangrove.

$$AGB = 0.112\left(\rho DBH^2 H\right)^{0.916} \quad (2)$$

where *AGB* is in kg, ρ is the wood specific density (oven-dried on a green volume basis, g cm^{-3}), *DBH* in cm, and *H* the total tree height (m). When known (from the literature), species wood specific density was applied. When a species was not identified, or wood specific density values were not available, an average specific wood density of 0.728 g cm^{-3} was applied, calculated from the data collected for all other trees sampled under the CarboVeg-GB project. Eq. (2) was developed using data from trees within the DBH range of 5 to 156 cm. Where trees were sampled outside of this range, the entire plot was discarded and not used in further analyses (see Section 2.3.3).

2.3.2.2. Mangrove trees. Mangroves are abundant in Guinea-Bissau, especially the species *Avicennia germinans* and *Rhizophora mangle*, although some individuals belonging to the species *Laguncularia racemosa* were also found during the forest inventory campaigns. A careful literature review indicated that the equations of Fromard et al. (1998) were well suited for estimating the AGB of *A. germinans* (Eq. 3, 42.0 cm maximum DBH) and *L. racemosa* (Eq. 4, 9.6 cm maximum DBH). However, for *R. mangle* the equation of Imbert and Rollet (1989) was used (Eq. 5, maximum DBH not available). Mangrove allometric relationships are reviewed also by Komiyama et al. (2008).

$$AGB = 0.140 DBH^{2.40} \quad (3)$$

$$AGB = 0.102 DBH^{2.50} \quad (4)$$

$$AGB = 0.178 DBH^{2.47} \quad (5)$$

where *AGB* is in kg and *DBH* in cm.

2.3.2.3. Palm trees. Palm trees are also abundant in the forests of Guinea-Bissau, with *Elaeis guineensis* (oil palm) being the dominant species, although some *Borassus aethiopum* individuals were also measured. A comprehensive search of allometric equations aimed at estimating the AGB of oil palm as a function of stem height was carried out. Some published equations were found (e.g., Khalid et al., 1999; Thenkabail et al., 2010), which were developed for plantations where the oil palm stem height is much lower than that of the palms measured in Guinea-Bissau. We assumed that it was preferable to use an equation where the stem height range was appropriate for our measurements, rather than use an equation that was developed for this specific species but in a completely different environment (i.e., plantations). Therefore, an equation developed based on palms measured in Bolivia (Eq. 6, 33.0 m maximum stem height), and presented in Pearson et al. (2005), was used. For palm trees, the height estimator (*H*) was the stem height as this has been shown to be the best estimator for palm *AGB* (e.g., Gehring et al., 2011; Thenkabail et al., 2010).

$$AGB = 6.666 + 12.826 H^{0.5} \ln(H) \quad (6)$$

where *AGB* is in kg and *H* in m.

2.3.2.4. Dead trees. To estimate the AGB of dead trees, a different approach was followed assuming that the only part of the tree left was its trunk. Therefore, the AGB of dead trees was estimated (conservatively) as the product of the volume of a truncated cone by the wood specific density (Eq. 7) (Pearson et al., 2007).

$$AGB = \left(\frac{1}{3}\right)\pi H\left(r_1^2 + r_2^2 + r_1 r_2\right)\rho \quad (7)$$

where *AGB* is in kg, r_1 is the dead tree base radius (m), r_2 the dead tree top radius (m), and ρ is in kg m^{-3}. As ρ for dead standing trees was unknown, the lowest wood specific density for the species measured during the CarboVeg-GB project (i.e., 328 kg m^{-3} for *Trema orientalis*) was used on the assumption that a dead standing tree should have a sufficiently high wood specific density to allow it to remain standing (e.g., Pearson et al., 2005).

2.3.3. Scaling-up AGB and data screening

From the tree-level estimates of *AGB*, stand-based estimates were generated by scaling these to the unit of 1 ha, and as a function of the nested plot size, such that (Eq. 8):

$$SF = 10{,}000 / A_{np} \quad (8)$$

Where *SF* is the scaling factor, and A_{np} is the nested plot area (m^2). As an example of the nest sizes adopted for this project (i.e., 4, 14, and 20 m radius), the SF for converting the plot data to a hectare basis ranged from 198.94 for the smallest, 16.24 for the intermediate, and 7.96 for the largest nested circular plot. Upscaling AGB to a reference area (i.e., hectare) was accomplished by combining individual measurements with the corresponding scaling factors (Eq. 9).

$$AGB = \sum_{i=1}^{n} agb_i sf_j \quad (9)$$

where *AGB* is in Mg ha^{-1}, agb_i is the plot *i*th tree AGB (Mg), and sf_i the scaling factor dependent on the *j*th plot size.

Of the 140 plots that were measured, 28 were discarded for several reasons. First, where plot locations were considered to be incorrectly placed or too close to non-paved roads or rivers, as determined through reference to Google Earth ™ high resolution images, these were omitted, as the ALOS PALSAR FBD backscatter intensity is influenced by these types of non-forest background (i.e., bare soil, water). In several cases, plots contained trees with DBH greater than the maximum used to build the equation of Chave et al. (2005) or orchard trees (e.g., mango, cashew) and/or bamboo that were not measured were also discarded. In the final analysis, 112 plots were used to establish relationships with ALOS PALSAR FBD data (Table 1). Table 1 displays also the AGB minimum, maximum, mean, and standard deviation per forest type for the screened AGB dataset.

3. Remote sensing data and methods

3.1. The ALOS PALSAR sensor

The Japan Aerospace Exploration Agency's (JAXA) launched ALOS on January 24, 2006, placing the satellite in a polar, sun synchronous orbit at ~700 km and ensuring a 46-day repeat cycle. PALSAR is one of the instruments onboard ALOS, which is an enhanced version of the Japanese Earth Resources Satellite (JERS-1) SAR instrument (Rosenqvist et al., 2007). PALSAR has a center frequency of 1270 MHz (23.6 cm, *i.e.*, L-band), and a chirp bandwidth of 14 MHz and 28 MHz. The instrument operated in five different observation modes, (i) Fine Beam Single (FBS), (ii) FBD, (iii) Polarimetric (PLR), (iv) ScanSAR, and (v) Direct Transmission (DT). However, data acquired in the FBD mode was used, with these providing HH/HV or VV/VH polarizations (14 MHz bandwidth). While 18 alternative off-nadir viewing angles were available (from 9.9° to 50.8°), only that acquired at 34.3° (HH and HV) polarization were selected by JAXA and used with these acquired in ascending node with

a 70 km swath width. Rosenqvist et al. (2007) provide a complete reference and description of the ALOS PALSAR sensor.

3.2. ALOS PALSAR data

3.2.1. Conversion from raw to single look complex (SLC) data

In support of the JAXA Kyoto and Carbon (K&C) Initiative and Principal Investigator (PI) programs (Rosenqvist et al., 2007), 26 ALOS PALSAR fine beam dual (FBD; HH and HV polarizations) scenes were acquired to cover the entire territory of Guinea-Bissau, between May 2 and June 5 2008. The data were acquired with an off-nadir angle of 34.3° (ascending orbits) (Fig. 1), and were provided by JAXA in level 1.0 (raw data). The Modular SAR Processor (MSP) developed by GAMMA Remote Sensing and Consulting AG (GAMMA) (http://www.gamma-rs.ch/) was used to process data from level 1.0 (raw) to 1.1 (SLC), and followed standard SAR processing (e.g., Oliver and Quegan, 2004; Werner et al., 2000; Woodhouse, 2005).

3.2.2. Conversion from SLC to geocoded terrain corrected (GTC) data

Prior to geocoding, the SLC data were converted to multi-look intensity (MLI) format. To obtain approximately square pixels in ground range coordinates, a multi-look factor of 1 in range and 5 in azimuth was used. The resulting MLI data was geocoded to a ground resolution of 12.5 m (both in range and azimuth). No filtering for speckle noise reduction was performed. The geocoding of the MLI data, which refers to a transformation from the slant-range/azimuth geometry to map projection geometry, was terrain corrected (GTC). GTC requires a digital elevation model (DEM) and was applied to the scenes acquired over Guinea-Bissau. A 90-m DEM mosaic retrieved from the Shuttle Radar Topography Mission (SRTM) over Guinea-Bissau was used. The 90-m DEM/SRTM mosaic was obtained from the Global Land Cover Facility (University of Maryland, USA), by concatenating the five Worldwide Reference System-2 (WRS-2) DEM/SRTM scenes covering Guinea-Bissau (path/row: 203/51, 203/52, 204/51, 204/52, and 205/51) (United States Geological Survey; USGS, 2006). These scenes were provided in a Universal Transverse Mercator (UTM) coordinate system (zone 28 N), WGS 84 datum, and were resampled to 50 m using bilinear convolution. The geocoding procedure was implemented using the Differential Interferometry and Geocoding (DIFF&GEO) module available from GAMMA. Refinement of the geocoding procedure can be undertaken by estimating local offsets between the actual SAR image and one simulated from the SRTM-derived DEM through an automated cross-correlation analysis. This refinement is generally successful over hilly areas, however, as much of the study area has very little topographic variation, the refinement was undertaken using Landsat 5 TM data acquired in 2007. The five WRS-2 Landsat 5 TM scenes covering Guinea-Bissau (path/row: 203/51, 203/52, 204/51, 204/52, and 205/51) were acquired from the USGS Global Visualization Viewer (GloVis), and spanned from April 1 to April 26, 2007 (Landsat 5 TM scenes from 2008 were not available from the USGS with lower cloud coverage). Only the Landsat 5 TM near-infrared band (band 4) was used to refine the geocoding of the ALOS PALSAR FBD data. This band was used as greater spatial variability is typical to this spectral region (Chavez, 1992), and, therefore, better cross-correlation with the ALOS PALSAR FBD data was anticipated. ALOS PALSAR FBD data geocoding errors were between 0.136 and 1.474 pixels (see Table 3), but most were well below 0.5 pixels. The scene with the higher mean offset in azimuth (ALPSRP124270200, 1.474) covers a few offshore islands belonging to Guinea-Bissau (lower left scene, Fig. 1). No field data was collected over that scene, so it was not considered an issue for the purpose of establishing a relationship between ALOS PALSAR FBD and field data.

Topographic normalization and incidence angle correction was undertaken using the procedure reported by Santoro et al. (2006), modified from Castel et al. (2001) (Eqs. 10 and 11), such that:

$$\sigma^0_f = \sigma^0_{cor}\left(\frac{\cos\theta_{iref}}{\cos\theta_{iloc}}\right)^n \tag{10}$$

$$\sigma^0_{cor} = \sigma^0\frac{A_{flat}}{A_{slope}} \tag{11}$$

where $\sigma^0{}_f$ represents the topographically normalized backscattering intensity (hereafter referred to as gamma nought, γ^0), σ^0 the uncorrected backscatter intensity (hereafter referred to as sigma nought, σ^0), θ_{iref} the incidence angle at a reference location (e.g., mid-swath)

Table 3
Information about the geocoding errors of each ALOS PALSAR FBD scene acquired over Guinea-Bissau.

Scene ID	Acquisition date (dd-mm-yyyy)	Center latitude (°)	Center longitude (°)	[a]Offsets	Range error (pixels)	Azimuth error (pixels)
ALPSRP121060210	02-05-2008	11.348	−14.249	18.3	0.221	0.149
ALPSRP121060220	02-05-2008	11.845	−14.349	23.6	0.208	0.138
ALPSRP121060230	02-05-2008	12.342	−14.449	7.8	0.307	0.159
ALPSRP121060240	02-05-2008	12.839	−14.548	0.7	0.286	0.149
ALPSRP121790200	07-05-2008	10.854	−15.761	1.9	0.323	0.230
ALPSRP121790210	07-05-2008	11.351	−15.861	5.1	0.234	0.317
ALPSRP121790220	07-05-2008	11.848	−15.961	5.2	0.353	0.212
ALPSRP121790230	07-05-2008	12.345	−16.060	9.1	0.370	0.177
ALPSRP122810230	14-05-2008	12.342	−13.378	2.6	0.638	0.196
ALPSRP122810240	14-05-2008	12.839	−13.478	0.9	0.494	0.209
ALPSRP123540200	19-05-2008	10.817	−14.682	3.6	0.201	0.136
ALPSRP123540210	19-05-2008	11.314	−14.781	15.4	0.211	0.139
ALPSRP123540220	19-05-2008	11.811	−14.881	15.9	0.246	0.152
ALPSRP123540230	19-05-2008	12.308	−14.981	3.9	0.417	0.185
ALPSRP123540240	19-05-2008	12.805	−15.081	0.9	0.477	0.169
ALPSRP124270200	24-05-2008	10.849	−16.299	1.9	0.324	1.474
ALPSRP124270210	24-05-2008	11.346	−16.399	4.2	0.340	0.571
ALPSRP124270220	24-05-2008	11.843	−16.498	2.6	0.392	0.383
ALPSRP124270230	24-05-2008	12.340	−16.598	4.3	0.410	0.301
ALPSRP125290220	31-05-2008	11.846	−13.817	18.9	0.222	0.148
ALPSRP125290230	31-05-2008	12.343	−13.918	7.2	0.513	0.165
ALPSRP125290240	31-05-2008	12.839	−14.017	0.3	0.463	0.225
ALPSRP126020200	05-06-2008	10.851	−15.223	1.6	0.312	0.216
ALPSRP126020210	05-06-2008	11.348	−15.322	5.1	0.409	0.224
ALPSRP126020220	05-06-2008	11.845	−15.422	4.5	0.387	0.202
ALPSRP126020230	05-06-2008	12.341	−15.523	3.8	0.375	0.206

[a] Percentage of cross-correlation windows where correlation between ALOS PALSAR and Landsat 5 TM data (near-infrared band) were above a given signal/noise ratio, and, therefore, were used to geocode each scene; 16,384 cross-correlation windows were tested in each scene.

and θ_{iloc}, the local incidence angle; σ^0_{cor}, the corrected backscattering coefficient, is obtained from the SAR pixels size for flat terrain (A_{flat}) and the true local SAR pixel size (A_{slope}). The coefficient n was set to 1 although can be adjusted according to the prevalent scattering processes as a function of land cover (e.g., forest/non-forest). In our case θ_{iref}, the incidence angle at mid-swath was set to 38.9°.

After each scene had been geocoded, all were combined to create a mosaic. First, the scenes corresponding to the same orbit (date) were combined to create a same orbit mosaic (i.e., a strip). Afterwards, the seven strip mosaics were combined to create the final ALOS PALSAR FBD mosaic over Guinea-Bissau. As the ALOS PALSAR sensor is a right-looking SAR, the basic rule to create this final mosaic was to allow that the near range (left hand-side of each strip) of each orbit mosaic to be placed on top of the far range (right hand-side of each strip) adjacent orbit mosaic, thus eliminating some signal degradation that might occur at the far range of the orbit (Lucas et al., 2010).

Conversion of ALOS PALSAR FBD backscatter intensity (γ^0) to dB scale was accomplished using Eq. (12).

$$\gamma^0_{dB} = 10\log_{10}\left(\gamma^0_I\right) \tag{12}$$

where $\gamma^0{}_{dB}$ is the ALOS PALSAR gamma nought backscatter intensity in dB scale, and $\gamma^0{}_I$ is the ALOS PALSAR gamma nought backscatter intensity in the original scale. The equation does not include the calibration factors for the HH and HV polarizations, as these were already included when processing from level 1.0 (raw) to level 1.1 (SLC). Fig. 3 shows the mosaic of the 26 ALOS PALSAR FBD geocoded scenes over the territory of Guinea-Bissau.

3.2.3. Extraction of ALOS PALSAR FBD data at field plot locations

Each plot measured in the field had a maximum radius of 20 m, corresponding to the external nested concentric sub-plot. Although the coordinates of each plot center were collected with Global Positioning System (GPS), there are always positional errors, especially in Guinea-Bissau where differential correction was unavailable (errors up to 8–10 m are common). The same applies to the ALOS PALSAR data (12.5 m ground pixel spacing), with geocoding errors well documented in the previous section (Table 3). Therefore, to compensate for these two sources of position errors, a buffer around each plot center with a 25-m radius was created. The buffer was selected as anything larger would have impinged on areas that were not homogeneous in terms of the plot vegetation type. All ALOS PALSAR pixels inside each 25-m buffer around each plot center were extracted, with several metrics computed (mean, minimum, maximum, and standard deviation), and used to establish relationships with the AGB at the plot level. As the original ALOS PALSAR FBD mosaic had a 12.5 m spatial resolution, and the buffer around each plot center was set to 25 m, then, the extracted values per plot were those located on a 4×4 pixel window size centered on each plot center.

To assess the amount of speckle of the processed ALOS PALSAR FBD MLI data (12.5 m ground resolution), the equivalent number of looks (ENL) (Eq. 13) (e.g., Oliver and Quegan, 2004; Woodhouse, 2005) was estimated over a set of twenty homogeneous areas belonging to the closed forest class.

$$\mathrm{ENL} = \mu^2/\sigma^2 \tag{13}$$

where μ and σ^2 are the mean and variance of the backscatter intensity values (original scale). The estimated ENL mean values were 2.42 and 2.37, for the HH and HV polarizations respectively. These values are approximately half of the number of looks used to produce the MLI data, meaning that the amount of speckle anticipated during multi-looking is lower than that estimated by Eq. (13). However, as mentioned above, the ALOS PALSAR FBD data extracted per plot were equivalent to extracting data located on a 4×4 pixel window size centered on each plot center. Therefore, the amount of speckle present in the ALOS PALSAR FBD data used to model AGB should be that of the ALOS PALSAR FBD data averaged to a 50-m spatial resolution. The ENL corresponding to this ALOS PALSAR FBD data were estimated, using the same set of twenty homogeneous areas, and the mean values were 17.02 and 16.85, for the HH and HV polarizations respectively. It shows that the amount of speckle was substantially reduced by spatially averaging the ALOS PALSAR FBD data.

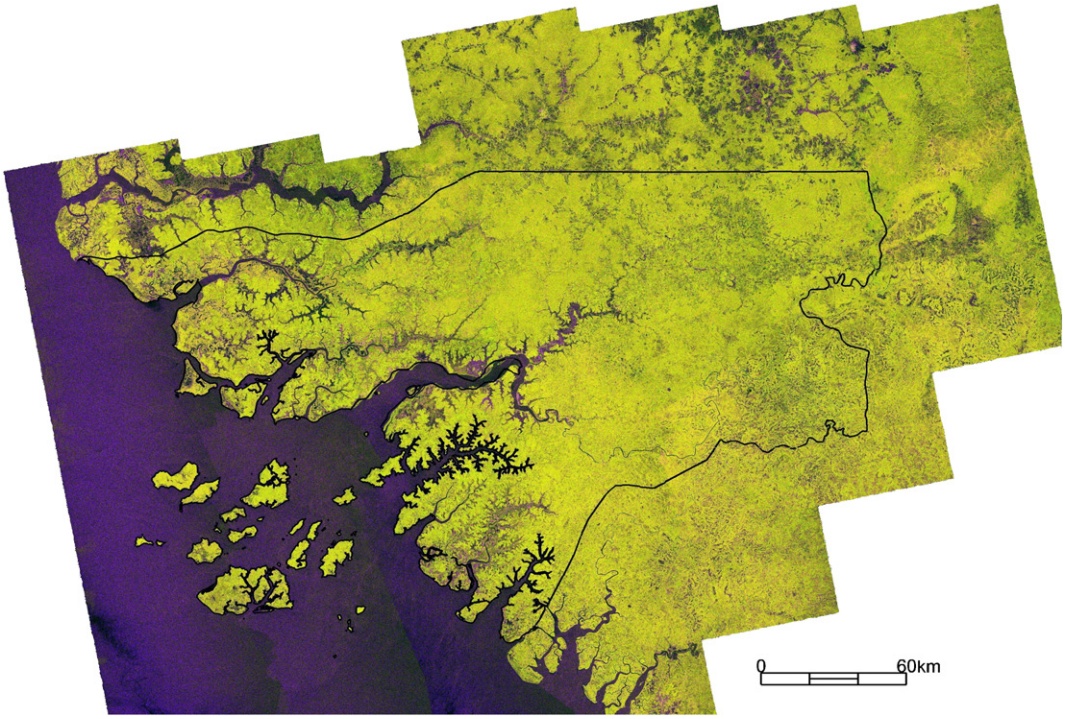

Fig. 3. Mosaic of the 26 ALOS PALSAR FBD geocoded scenes covering Guinea-Bissau (country limit outlined in black). RGB image, where R: HH polarization (dB scale), G: HV polarization (dB scale), and B: HH divided by HV (original scale).

3.3. Modeling the relationship between ALOS PALSAR FBD intensity backscatter and AGB

3.3.1. Contribution of different ALOS PALSAR polarizations

A model (Eq. 14) was fitted to evaluate the strength of the relationship between ALOS PALSAR backscatter intensity (γ^0) and AGB data. Lucas et al. (2006) have identified this type of model as characteristic of the relationship between SAR and forest biophysical parameters in sub-tropical savannas of Australia (Queensland). The proportion of variance explained by the model (similar to the coefficient of determination for linear regression models) was used as an indicative measure of model adjustment and, therefore, relationship strength.

$$\gamma^0 = \beta_s + \left((\beta_n - \beta_s)e^{-kAGB}\right) \quad (14)$$

where γ^0 is the corrected (topographic normalization and incidence angle correction) ALOS PALSAR backscatter intensity data (HH or HV, dB scale), *AGB* is the aboveground biomass (Mg ha^{-1}), β_n the mean backscatter value (HH or HV) of non-forest plots, and β_s and k the parameters to be estimated. The calculated β_n values were: HH = −11.66 dB, and HV = −18.85 dB.

Saatchi et al. (2007) focused on estimating the AGB of components from single-frequency or/and multi-frequency polarimetric SAR data based on relationships established with forest structure and biomass components. The algorithm considered scattering (e.g., direct volume and volume surface or double bounce) as a result of the differential interaction of microwaves of varying frequency and polarization with the live stems, branches and foliage and as a function of their size and density (which relate to biomass) and moisture content. The equations used considered the relative area filled by the forest canopy and the transmissivity of the vegetation as a function of biomass, with this reflecting the loss of sensitivity of radar backscatter, particularly at higher levels of AGB.

An equation, derived from Saatchi et al. (2007), was used to model AGB as a function of ALOS PALSAR FBD intensity data (Eq. 15):

$$\ln(\mathrm{AGB}) = a_0 + a_1\gamma^0_{HH} + a_2{\gamma^0_{HH}}^2 + a_3\gamma^0_{HV} + a_4{\gamma^0_{HV}}^2 \quad (15)$$

where ln*(AGB)* refers to the natural logarithm of the AGB, $\gamma^0{}_{HH}$ is the to HH backscatter intensity (γ^0, dB), and $\gamma^0{}_{HV}$ the HV backscatter intensity (γ^0, dB). The model was fitted using an ordinary least squares estimator and assessed using a 10-fold cross-validation approach. To obtain estimates of AGB in the original scale, the anti-logarithmic function was applied to the AGB data in the logarithmic scale.

3.3.2. A machine learning approach (bagging stochastic gradient boosting, BagSGB)

Traditionally, the attempt of explaining a given variable (dependent variable) as a function of one or more variables (predictor or independent variable(s)) has relied on parametric statistical models, such as linear (simple or multiple) regression models. These models have several assumptions, including normal distribution of errors and variables, as well as homoscedasticity (e.g., Steel and Torrie, 1980). However, in the past decades, and largely because of problems presented by large arrays of data, several other methods and algorithms have been developed, mainly from computer theory, which are commonly referred to as machine learning methods (e.g., Witten and Hall, 2011). These methods have a large scope of application, from prediction (e.g., Elith et al., 2008; Moisen et al., 2006) to classification (e.g., Carreiras et al., 2006; Friedl et al., 2010) problems, and can have different formulations (e.g., neural nets, ensembles of trees, support vector machines).

The algorithm being tested in this study (SGB; Friedman, 2001, 2002) originated from the decision tree theory (e.g., Breiman et al., 1984; Hastie et al., 2001). Decision tree theory relies on partitions of the space of all possible predictor variables. Starting with the whole predictor space (at the root of the tree), the space is successively split using a series of rules such that, in the end, each terminal node of the tree is assigned to the most probable response class (classification trees) or the mean response in that node (regression trees) (Breiman et al., 1984; Safavian and Landgrebe, 1991). Some advantages over traditional parametric classification methods include the non-parametric nature of the classifier, quantification of variable importance, disclosure of non-linear and hierarchical relationships between predictor variables, and acceptance of missing values (Breiman et al., 1984; Friedl and Brodley, 1997; Hansen et al., 1996).

However, classification and regression trees are sensitive to small perturbations in the training data, which may originate large changes in the resulting outputs (Breiman, 1996). Therefore, these unstable methods can have their accuracy improved with a perturbing and combining technique. This generates multiple perturbed versions of the classifier (a.k.a., ensemble or committee) and combines these into a single predictor (Breiman, 1998). These methods can be divided into two types: those that adaptively change the distribution of the training set based on the performance of previous classifiers (e.g., boosting) and those that do not (e.g., bagging) (Bauer and Kohavi, 1999).

SGB (Friedman, 2001, 2002) combines both the advantages of bagging and boosting and can be used in regression and classification problems (e.g., Elith et al., 2008; Moisen et al., 2006). It typically uses a base learner (in our case, decision trees) and constructs additive regression (or classification) models by sequentially fitting the chosen base learner to current "pseudo"-residuals by least squares at each iteration (Friedman, 2002). At these iterations, a simple base learner is built using a random sub-sample of the training data (without replacement), which has been shown to substantially improve the prediction accuracy and, execution speed (computational saving), and makes the approach resilient to overfitting (Friedman, 2002). Also, Friedman (2001) developed an improved measure of variable importance for boosting estimates, which is to average the relative influence of a given explanatory variable across all trees generated by the boosting algorithm (Ridgeway, 2007); this is a helpful parameter that is common to all decision tree algorithms and allows the disclosure and ranking of predictive variables. The final model is a linear combination of each simple base learner, which can be seen as a regression model whereby each term is a tree (Elith et al., 2008). SGB have been used in several problems aimed at predicting or classifying spatial data (e.g., De'ath, 2007; Elith et al., 2008; Jalabert et al., 2010; Lawrence et al., 2004; Moisen et al., 2006; Sankaran et al., 2008). Furthermore, some studies (e.g., Ganjisaffar et al., 2011; Suen et al., 2005) have demonstrated that building and combining (by averaging in the case of regression) several SGB models on bootstrap samples of the training dataset performs significantly better than an unique SGB model. Particularly, Suen et al. (2005) showed that bagging SGB models had a better performance in terms of regression modeling than SGB itself over a collection of real world datasets, and concluded that it was accomplished by variance reduction. Therefore, instead of building a single SGB model, we built several SGB models fitted to bootstrap samples (with replacement) of the original training dataset (n = 112). The final model (bagging SGB, or BagSGB) was built by averaging the predictions from the several SGB models fitted to the bootstrap samples. In this study we used 25 bootstrap replicates to build a BagSGB model, each bootstrap replicate evaluated with a 10-fold cross-validation approach. Breiman (1996) suggests that a higher number of replicates tend not to produce a significant test set error reduction. Also, as each predicted observation is the result of averaging over a set of 25 bootstrap replicates, it allows also building a measure of prediction variability. The coefficient of variation (CV), calculated as the standard deviation of a predicted observation divided by the corresponding average value, was used as a measure of assessing the uncertainty associated with that predicted observation.

We used the R (R Development Core Team, 2011) gbm package (Ridgeway, 2007), Elith et al. (2008) R code, as well as in-house R

code to implement SGB. Model fitting under SGB has a number of options including, (i) distribution (e.g., Gaussian, Laplace), (ii) tree complexity (i.e., the number of nodes in each decision tree), bagging fraction (i.e., the random fraction of the training data used to build each decision tree), shrinkage rate (i.e., controlling the learning rate of the algorithm), and training fraction. In our case, using Elith et al. (2008) R code, a 10-fold cross-validation approach was used to define the optimal number of iterations in each SGB model fitting.

In-house R code was developed to implement a loop using Ridgeway (2007) and Elith et al. (2008) R code, so that a set of SGB options could be tested for building each model from each of the bootstrap samples. Those were, (i) Gaussian distribution (i.e., the loss function, whose measure of deviance to be minimized is the sum of the squared error (Ridgeway, 2007)), (ii) bagging fraction = 0.5, 0.6, 0.7, 0.8, and 0.9; (iii) tree complexity = 1, 2, 3, 4, 5, 7, 9; and (iv) learning rate = 0.01, 0.005, 0.0025, 0.001, and 0.0005. Therefore, the selection of the best SGB model fitted to a given bootstrap sample is the result of evaluating 175 candidate individual SGB models. In the end, 4375 SGB models were evaluated for the 25 bootstrap samples. Useful information regarding SGB applied to biophysical and ecological modeling can be found in, for example, De'ath (2007), Elith et al. (2008), and Moisen et al. (2006). Elith et al. (2008) refer that small learning rates result in a smaller contribution of each tree, and, therefore, are more directed at building a model that will provide a more consistent estimate of the dependent variable; as a rule of thumb, they indicate that, as a minimum, models should be fitted with 1000 trees. Accordingly, to select the best SGB model out of the various candidates for a given bootstrap sample, we selected the model that had the lowest cross-validated deviance (i.e., lowest root mean square error), given that the number of trees producing that model were not below 1000.

The previous loop was applied to the various SGB models, where the AGB was chosen as the dependent variable, and the mean, minimum, maximum, and standard deviation of the ALOS PALSAR HH and HV backscatter intensity values extracted for each plot (25 m buffer around each plot center) were selected as independent variables. The number of pixels that were used to compute those metrics ranged from 11 to 14, the variation being dependent on the plot location regarding the 12.5-m ALOS PALSAR mosaic.

3.3.3. Model comparison and evaluation

The fitted models (semi-empirical and SGB) were compared using a traditional bias and variance decomposition of the root mean square error (Eqs. 16, 17, and 18) (e.g., Hastie et al., 2001):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} = \sqrt{\sigma^2 + b^2} \tag{16}$$

where $RMSE$ is the model root mean square error, e_i the error (difference between the observed and predicted values) of the ith observation, σ^2 the error variance, and b the error bias;

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^2 \tag{17}$$

where n is the number of observations, and $\bar{e}$ the mean error;

$$b = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{18}$$

Model evaluation was achieved by a 10-fold cross-validation approach (e.g., Hastie et al., 2001), as the number of observations was not large enough to evaluate model performance with an independent subset. The correlation coefficient between observed and predicted (10-fold cross-validation) AGB were used to assess model predictability.

3.4. Estimation of carbon stocks and comparison with published biomass maps

Average and total AGB and carbon stocks per land cover class (i.e., closed forest, open forest, savanna woodland, and mangrove) were estimated, as well as the total AGB and carbon stock in the forests of Guinea-Bissau. A conversion factor of 0.5 was used to convert AGB to carbon content (e.g., Saatchi et al., 2011a; Silva et al., 2011).

However, a comparison with published biomass maps, especially when derived from distinct data sources is important, namely to compare and assess differences in AGB and carbon stocks at a regional and/or national level. Two published maps were selected, Baccini et al. (2008) and Saatchi et al. (2011a). Baccini et al. (2008) mapped AGB across tropical Africa (1-km spatial resolution) using an ensemble of regression tree-based models (random forests) that related *in situ* measurements with data acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard NASA's Terra and Aqua satellites between 2000 and 2003. A cross-validation approach showed that the model explained 82% of the variance in AGB, corresponding to a RMSE of 50.5 Mg ha^{-1} (for AGB values between 0 and 454 Mg ha^{-1}). Mitchard et al. (2011) assessed the AGB map of Baccini et al. (2008) using an independent dataset of 1154 field measurements obtained in 16 African countries, and concluded that large errors were associated with the Baccini et al. (2008) AGB map, especially large underestimation in forests and woodland savannas, resulting in a RMSE of 145 Mg ha^{-1}. Saatchi et al. (2011a) mapped forest carbon stocks [AGB plus below-ground biomass (BGB)] in the pan-tropical belt (1-km spatial resolution) around the year 2000 using a data fusion algorithm based on the maximum entropy approach (MaxEnt) to model field-based measurements as a function of data acquired by the Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and land Elevation Satellite (ICESat), and extrapolating to the landscape using data from other optical and microwave sensors. The prediction variability ranged from ±6% to ±53% on a pixel basis, but when scaling-up to project- and country-scales, the errors decreased to ±5% and ±1% respectively. A useful component of the Saatchi et al. (2011a) work was the production of uncertainty estimate on a pixel-by-pixel basis.

Furthermore, another approach for comparing the results from this work and the two abovementioned published maps was followed. To compare mean AGB and AGB carbon stock estimates obtained from the two published maps and the work presented here, a systematic square grid was placed over the three maps and the total AGB carbon stock (kt) and average AGB (Mg ha^{-1}) was estimated per grid cell. As our map has a 50-m spatial resolution and the two published datasets have a spatial resolution of around 1-km, we chose a 10-km spatial resolution systematic grid. Only those 10-km grid cells where more than 75% of the area was covered by forests (according to the land cover map of 2007 mentioned in Section 2.2) were considered, as values lower than that were not considered valid for comparison.

4. Results and discussion

4.1. Relationship between ALOS PALSAR FBD backscatter intensity and AGB data

The relationship between AGB and the ALOS PALSAR FBD mean backscatter intensity (individual HH and HV polarizations, γ^0) are illustrated in Fig. 4. The proportion of variance explained by the fitted equations is low. Furthermore, the strongest relationships were observed when using HV polarization data which aligns with previous findings (e.g., Lucas et al., 2006; Mitchard et al., 2009). The most probable cause for the lack of relationship between ALOS PALSAR individual polarization backscatter intensity and AGB is the complex structure (both vertically and spatially) of the Guinea-Bissau forest vegetation (see Fig. 2), with many different species, including palm trees, occurring

on the same plot, and sharply varying values of DBH and height. It has also been shown by other studies that the size of the sampled plot has influence on the relationship strength (e.g., Mitchard et al., 2009; Saatchi et al., 2011b). As examples, Mitchard et al. (2009) studied the relationship between ALOS PALSAR data (FBD, 34.3° off-nadir angle) and AGB across four African landscapes. They obtained a good relationship ($R^2 = 0.73$) using a logarithmic model that related ALOS PALSAR HV intensity backscatter to AGB. However, the field plot data they used were taken from plots of different size (between 0.07 ha and 1.86 ha), with most (around 80%) being greater than 0.2 ha (which is almost twice the size of the outer nested plot in our study, 20-m radius). In a simulation study, Carreiras et al. (Unpublished results) has shown that increasing the plot size results in a better relationship between ALOS PALSAR backscatter intensity (especially with the HV polarization) and AGB in sub-tropical savannas of Australia (Queensland). Therefore, and especially in ecosystems where biodiversity is high, and the structural vegetation types are complex, it seems that a better relationship between L-band SAR data and AGB can be obtained if biomass values are measured over larger areas.

There is also an AGB saturation level (around 100 Mg ha^{-1}) (Fig. 4) above which the ALOS PALSAR backscatter intensity does not increase with an increase in AGB. Saturation values for L-band SAR data similar to this have also been referred by several authors (e.g., Austin et al., 2003; Dobson et al., 1992; Fransson and Israelsson, 1999; Hoekman and Quinones, 2000; Imhoff, 1995; Lucas et al., 2006; Luckman et al., 1997; Saatchi et al., 2007).

4.2. Modeling AGB

4.2.1. Semi-empirical model

Eq. (15) was applied to the relationship between AGB estimated in the field and the ALOS PALSAR FBD backscatter intensity data. Table 4 shows the parameters estimates, fitting statistics, and the 10-fold cross-validation statistics for the tested model. Additionally, Fig. 5 shows the scatterplots of the observed versus predicted AGB, either using fitted or cross-validation values. Although there is an improvement of model fitting when using a model that incorporates both the HH and HV polarizations rather than the single case, the fitting statistics are also low ($R^2 = 0.27$, RMSE = 66.46 Mg ha^{-1}). The same pattern is observed in the 10-fold cross-validation statistics, with a high RMSE (70.06 Mg ha^{-1}) and low coefficient of correlation (0.23) between the observed and predicted AGB. These low correlations between observed and fitted/predicted statistics are further highlighted in Fig. 5 and are not close to the perfect agreement line. The larger contribution for the RMSE when using Eq. (16) came from the error variance, as the error bias is smaller in terms of magnitude. The model presented in Saatchi et al. (2007) is a semi-empirical algorithm and was tested in the Yellowstone National Park where coniferous forests, primarily dominated by lodgepole pine (*Pinus contorta* var. *latifolia*), make up around 80% of the forested area. The model was also derived from a physically-based model (i.e., Durden et al., 1989) and therefore several assumptions were necessarily made to describe the forest in a way that was simple enough to be entered into a parametric form (e.g., Moghaddam and Saatchi, 1999; Saatchi and Moghaddam, 2000). By contrast, the forests and savannas of Guinea-Bissau are very complex containing a diversity of structural forms as a function of species type and growth stage, with mixed evergreen and deciduous hardwood species dominating uneven-aged forest types of irregular structure. Also, the original model of Saatchi et al. (2007) developed the methods using fully polarimetric airborne SAR data whereas only the HH and HV polarization data were available for this study.

4.2.2. SGB modeling

The final BagSGB model was the result of combining the 25 SGB models built from the corresponding bootstrap samples of the original training dataset. The AGB values predicted from the 10-fold cross-validation of each of the 25 models were aggregated (averaged) and the corresponding statistics calculated (Table 5). The comparison between observed and cross-validation predicted AGB values is displayed in Fig. 6a. Each cross-validation predicted AGB value was the result of averaging between 15 and 47 times, as each bootstrap sample was built with replacement. A close correspondence (R = 0.95) between observed and cross-validation predicted AGB values was identified, although the AGB above 150 Mg ha^{-1} was underestimated. As a comparison, the same combination of parameters (i.e., distribution, bag fraction, tree complexity, and learning rate) was tested to build the best SGB model fitted to the original training dataset. This model was based on 2500 trees, resulting from a bag fraction of 0.6, a learning rate of 0.0005, and a tree complexity of 1. The 10-fold cross-validation results performed less well than those resulting from the application of an ensemble of SGB models (i.e., a BagSGB model). The cross-validation RMSE of this model was 62.84 Mg ha^{-1}(26.62 Mg ha^{-1} in the BagSGB model), the variance 3948.25 (708.55 in the BagSGB model), the error bias −0.47 Mg ha^{-1} (−0.57 Mg ha^{-1} in the BagSGB model), and the correlation (R) between observed and cross-validation

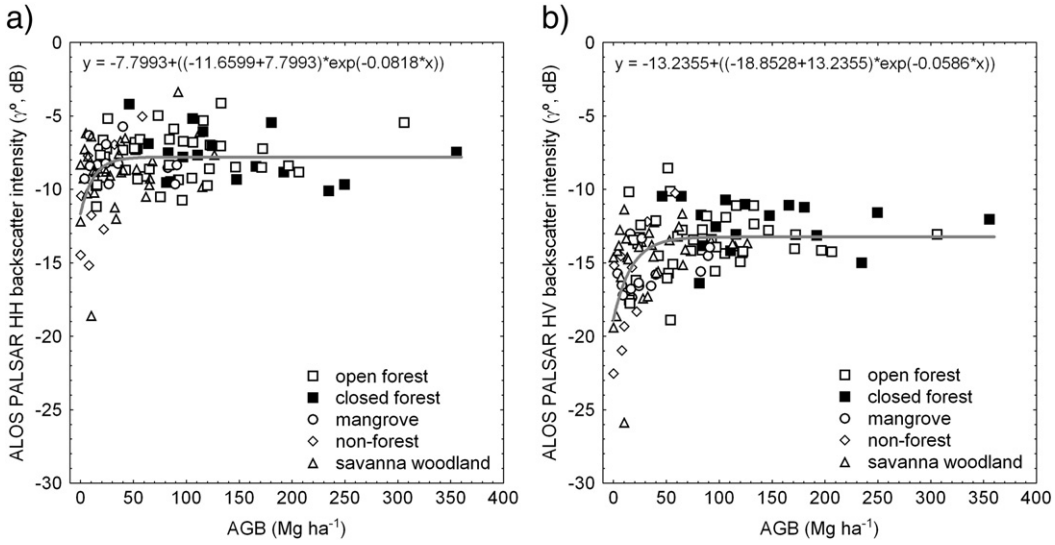

Fig. 4. Relationship between AGB (Mg ha^{-1}) and ALOS PALSAR FBD backscatter intensity: a) HH (γ^0, dB) (proportion of variance explained = 0.12); b) HV (γ^0, dB) (proportion of variance explained = 0.20). The corresponding fitted equations (Eq. 14) are also shown in each figure.

Table 4
Parameters estimates and fitting statistics, using the semi-empirical model displayed in Eq. (15).

Parameters estimates				
a_0	a_1	a_2	a_3	a_4
3.574	0.734[a]	0.043	−0.714	−0.034
Fitting statistics				
[b]*RMSE*	[c]s^2	[d]*b*	[e]R	[f]R^2
66.46	3882.70	23.13	0.33	0.27
10-fold cross-validation statistics				
[b]*RMSE*	[c]s^2	[d]*b*	[e]R	
70.06	4469.88	20.93	0.23	

[a] significant for an alpha level equal to 0.05.
[b] Root mean square error (Mg ha^{-1}).
[c] Error variance.
[d] Error bias (Mg ha^{-1}).
[e] Coefficient of correlation between observed and fitted values
[f] Coefficient of determination.

Table 5
10-fold cross-validation statistics for the fitted BagSGB model.

[1]*RMSE*	[2]s^2	[3]*b*	[4]R
26.62	708.55	−0.57	0.95

[1] Root mean square error.
[2] Error variance.
[3] Error bias.
[4] Correlation between observed and predicted values.

predicted AGB values was 0.38 (0.95 in the BagSGB model). Fig. 6b shows the scatterplot between observed and cross-validation predicted AGB values for this SGB model. The error bias is approximately the same, therefore, the RMSE decrease (from 62.84 to 26.62 Mg ha^{-1}) was a consequence of variance reduction from 3963.96 in the SGB model to 708.55 in the BagSGB model. This is in agreement with the results of Suen et al. (2005) already mentioned in Section 3.3.2.

To evaluate the degree of prediction variability associated with the fitted BagSGB model, the coefficient of variation of the AGB was calculated, using the average and standard deviation values of AGB obtained during the 10-fold cross-validation of this fitted model. The relationship between the 10-fold cross-validation predicted AGB and corresponding coefficient of variation values is shown in Fig. 7. The coefficient of variation is higher for lower AGB values, decreased until AGB values around 100 Mg ha^{-1}, and then increased again. However, these higher values of the AGB coefficient of variation are not that evident in the relationship between observed and cross-validation predicted AGB values (Fig. 6a), thus suggesting that the averaging process inherent to a BagSGB model results in a reduced variance. Nevertheless, we acknowledge that some bias may have been introduced in the fitted BagSGB model as a consequence of using bootstrap samples of the original training dataset. The bootstrap samples were built with replacement, so each bootstrap sample will have at least duplicates of some observations (i.e., plots). As the evaluation of the BagSGB model was performed by a 10-fold cross-validation approach, one cannot guarantee that the same observation is not present in the training subset (9/10 folds) and in the testing subset (1/10 fold), hence the cross-validation statistics may be skewed.

The cross-validation statistics of the fitted BagSGB model represent an improvement on those associated with the model fitted with Eq. (15) (i.e., AGB as a function of ALOS PALSAR FBD mean backscatter intensity values). The RMSE associated with the AGB estimation may also be reduced by using other metrics (e.g., minimum, maximum, and standard deviation). The RMSE decreased from 70.06 Mg ha^{-1} to 26.62 Mg ha^{-1}, the error variance decreased from 4469.88 to 708.55, the error bias decreased from 20.93 Mg ha^{-1} to −0.57 Mg ha^{-1}, and the correlation coefficient between observed and predicted values increased from 0.23 to 0.95. As mentioned earlier, the higher contribution to the RMSE came from the error variance and a more complex model approach (i.e., BagSGB) is more successful at predicting AGB than the simpler algorithm (Eq. 15), although this was adapted from a physically-based model.

The importance of each independent variable for the fitted BagSGB model is displayed in Fig. 8. The score of each variable is the result of averaging the scores among the 25 SGB models which composed the final BagSGB model. The most important variable is the mean value of the ALOS PALSAR FBD HV backscatter intensity, followed by the minimum value of the ALOS PALSAR FBD HV backscatter intensity, the minimum value of the ALOS PALSAR FBD HH backscatter intensity, and the standard deviation of the ALOS PALSAR FBD HV backscatter intensity. The other tested variables have lesser importance for the fitted SGB model. The standard deviation of the variable importance index for each metric is rather high, suggesting that the metrics used to build the 25 SGB models that compose the final BagSGB

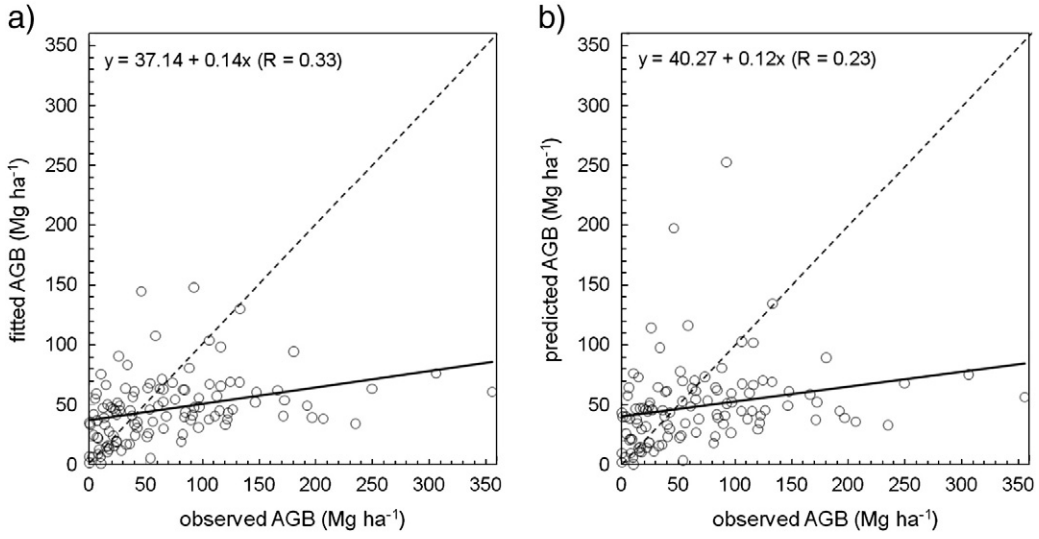

Fig. 5. Relationship between observed and predicted AGB, as a result of fitting Eq. (15); a) observed vs. fitted AGB; b) observed vs. 10-fold cross-validation AGB. The solid line represents the linear fit between observed and predicted values (the corresponding equation and coefficient of correlation are also shown); the dashed line represents what would be a perfect agreement relationship.

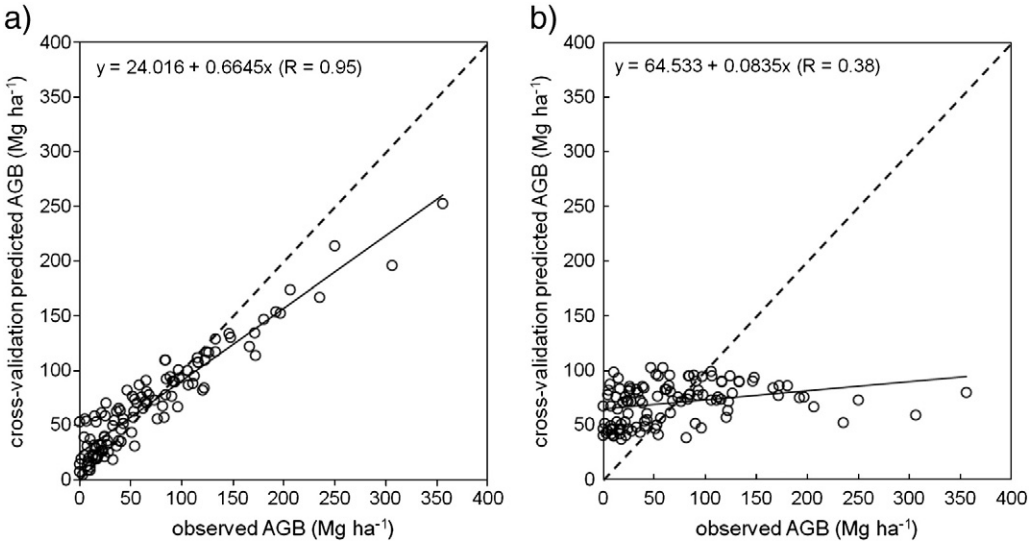

Fig. 6. Relationship between observed and cross-validation predicted AGB values, resulting from a) fitting a BagSGB model, and b) fitting a unique SGB model to the training dataset. The solid line represents the linear fit between observed and predicted values (the corresponding equation and coefficient of correlation are also shown); the dashed line represents what would be a perfect agreement relationship.

model could have had different weights, depending of the data structure (i.e., the observations that were randomly selected with replacement for each bootstrap sample).

On the basis of its better performance in the cross-validation assessment, the BagSGB was applied to the entire territory of Guinea-Bissau. The ALOS PALSAR FBD backscatter intensity data used to build the model were extracted in a 25 meter buffer around each plot center with this being equivalent to a spatial resolution of approximately 50 m. Therefore, to create the AGB map of Guinea-Bissau (Fig. 9), the original ALOS PALSAR FBD backscatter intensity data at 12.5 meter spatial resolution was necessarily aggregated to a 50 m spatial resolution and the minimum, maximum, mean and standard deviation values were computed. A map displaying the prediction variability of each pixel was also produced by computing the coefficient of variation (%) on a pixel-by-pixel basis (Fig. 10).

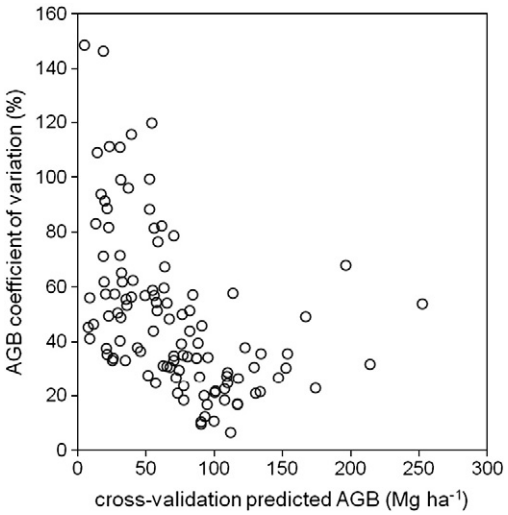

Fig. 7. Relationship between cross-validation predicted AGB values and the AGB coefficient of variation, resulting from fitting a BagSGB model.

4.3. Estimation of carbon stocks and comparison with published biomass maps

Total AGB and derived carbon stocks, as well as mean AGB and prediction variability, per forest type in Guinea-Bissau were calculated on the basis of the AGB map produced in this study as well as from the work of Baccini et al. (2008) and Saatchi et al. (2011a) (Table 6). Differences in the time frame of each study [2008 for this study, 2000–2003 for Baccini et al. (2008), and early 2000s for Saatchi et al. (2011a)] were not considered nor addressed here. The prediction variability at the national level (E_N) was estimated using the equation of Saatchi et al. (2011a) (Eq. 19):

$$E_N = \frac{\sqrt{\sum_{i=1}^{N}(AGB_i E_i)^2}}{\sum_{i=1}^{N} AGB_i} \quad (19)$$

where AGB_i and E_i are the AGB and prediction variability at the ith pixel respectively and N the number of pixels in the area being

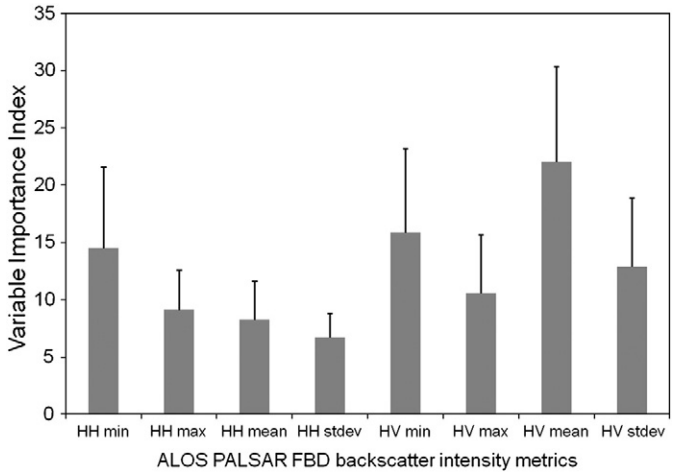

Fig. 8. Variable importance index of each metric for the fitted BagSGB; min—minimum; max—maximum; stdev—standard deviation. The standard deviation of the variable importance index for each metric is shown on top of each bar.

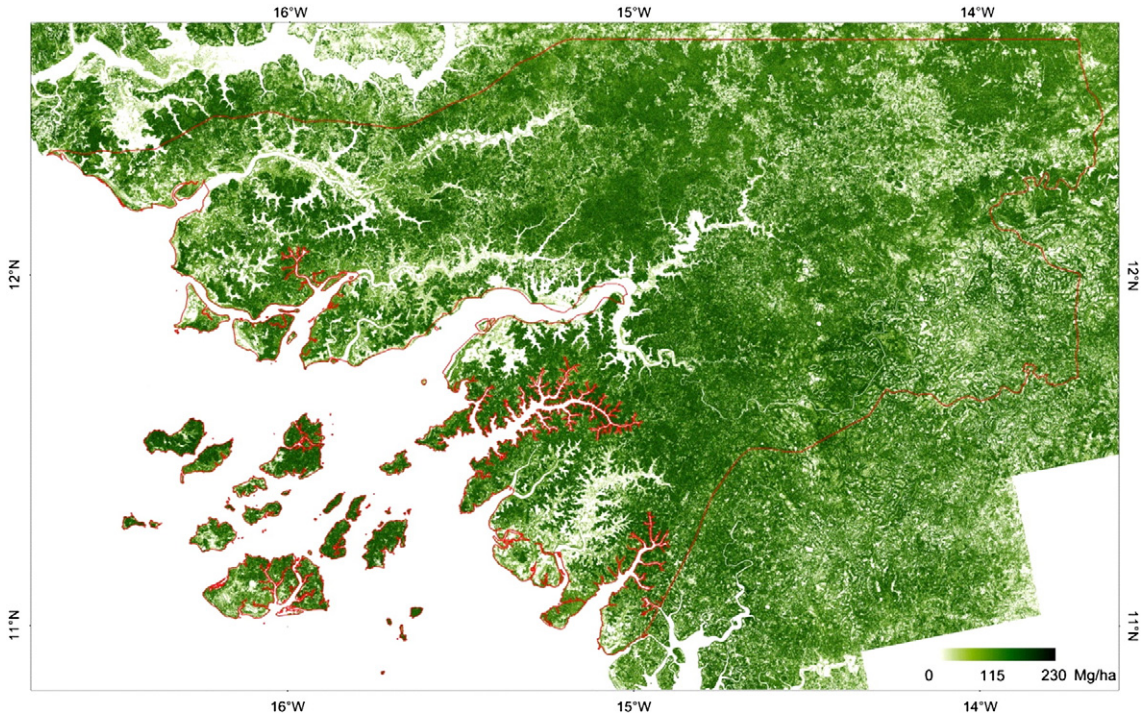

Fig. 9. Guinea-Bissau forest AGB map produced with a BagSGB model. The color ramp indicates the gradient of AGB values, from 0 to 230 Mg ha^{-1}, and the red line the country boundary.

assessed (in this case, the country). The AGB carbon stock derived from this study (96.93 Mt C) is closer to the value obtained from Saatchi et al. (2011a) (83.52 Mt C). The value estimated from Baccini et al. (2008) is approximately half of that estimated from this study (44.24 Mt C). The behavior is the same at the forest class level, with the AGB carbon stocks obtained from this study being closer to those of Saatchi et al. (2011a), the exception being the mangrove class where the value reported by this study (9.08 Mt C) was closer to that reported by Baccini et al. (2008; i.e., 8.44 Mt C) than Saatchi et al. (2011a; 6.97 Mt C). The overall and class-specific mean AGB values display the same pattern as the AGB carbon stock (i.e., the values reported in this study are closer to the ones obtained from Saatchi et al. (2011a) and higher than those reported by Baccini et al. (2008)). The exceptions are the mean AGB in the closed forest class, where the value obtained from Baccini et al. (2008) (153.04 Mg ha^{-1}) is higher than that reported in this study and obtained from Saatchi et al. (2011a)

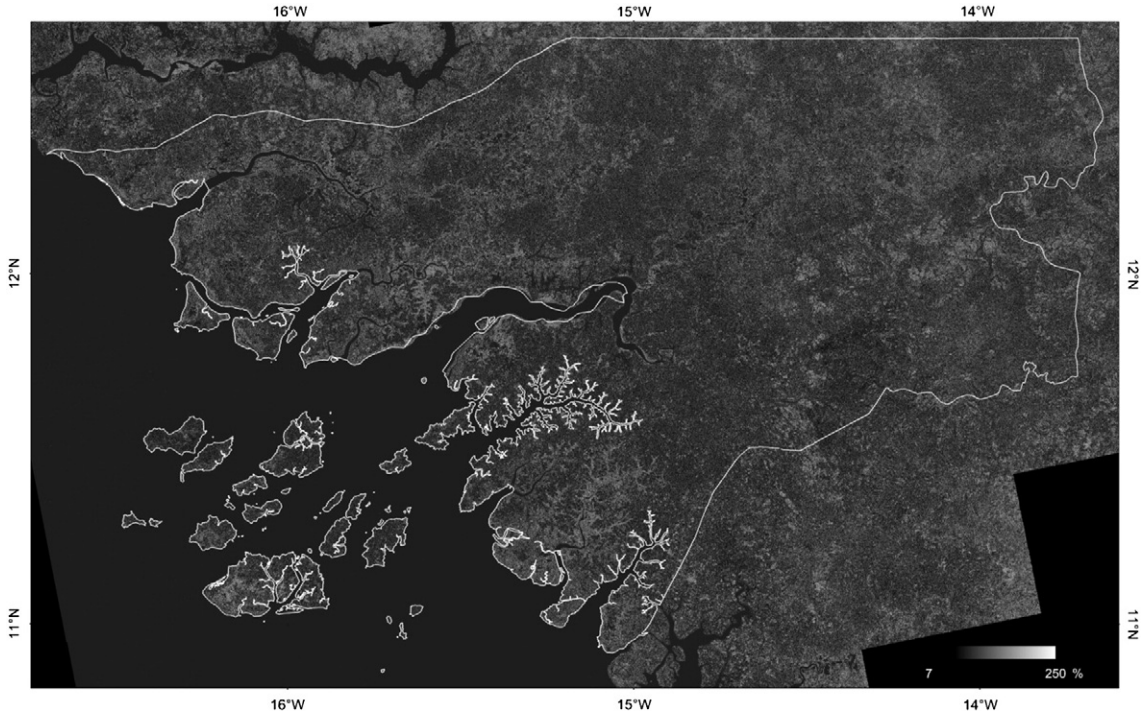

Fig. 10. Guinea-Bissau forest AGB prediction variability produced with a BagSGB model. The color ramp indicates the gradient of forest AGB coefficient of variation values, from 7 to 250%, and the white line the country boundary.

Table 6
Total AGB, AGB carbon stock, mean AGB, and prediction variability per forest type in Guinea-Bissau, derived from this study (BagSGB model), Baccini et al. (2008), and Saatchi et al. (2011a).

Forest classes	Area (ha)	Total AGB (Mt)	AGB carbon stock (Mt C)	Mean AGB (Mg ha^{-1})	Coefficient of variation (%)
This study					
Closed forest	43, 606	3.81	1.90	87.30	0.09
Open forest	839, 381	65.97	32.98	78.59	0.02
Savanna woodland	1, 768, 547	105.93	52.97	59.90	0.02
Mangrove	323, 422	18.16	9.08	56.15	0.04
Total	2, 974, 956	193.87	96.93	65.17	0.01
Baccini et al. (2008)					
Closed forest	43, 606	6.67	3.34	153.04	na
Open forest	839, 381	34.46	17.23	41.06	na
Savanna woodland	1, 768, 547	30.47	15.23	17.23	na
Mangrove	323, 422	16.88	8.44	52.19	na
Total	2, 974, 956	88.48	44.24	28.45	na
Saatchi et al. (2011a)					Error (%)
Closed forest	43, 606	4.14	2.07	94.98	2.84
Open forest	839, 381	58.67	29.33	69.89	0.52
Savanna woodland	1, 768, 547	90.29	45.14	51.05	0.25
Mangrove	323, 422	13.94	6.97	43.10	0.91
Total	2, 974, 956	167.04	83.52	55.68	0.24

(87.30 Mg ha^{-1} and 94.98 Mg ha^{-1} respectively). The mean AGB in the mangrove class (56.15 Mg ha^{-1}) was closer to that of Baccini et al. (2008; 52.19 Mg ha^{-1}). The measure of prediction variability at the national level, the coefficient of variation (%) in this study and the error (%) in Saatchi et al. (2011a) were low, with the highest values reported for the closed forest class (0.09% and 2.84% respectively). The values of prediction variability on a pixel basis ranged from 7–250% (mean = 42%) to 23−922% (mean = 30%) in this study and in Saatchi et al. (2011a) respectively. Although higher on a pixel level, the values decrease substantially when a measure of prediction variability is obtained at a regional- or national level, which has also been confirmed by Saatchi et al. (2011a).

A more detailed comparison between the results of this study and those from Baccini et al. (2008) and Saatchi et al. (2011a) was carried out using estimates of AGB carbon stocks and mean AGB values obtained from a grid with a 10-km spatial resolution (Fig. 11). There was a better agreement between the results from our study and Saatchi et al. (2011a) (Fig. 11c and 11d) than with the study of Baccini et al. (2008) (Fig. 11a and 11b). Therefore, for lower values of AGB carbon stocks and mean AGB values, our study produced estimates that were higher than those obtained from Saatchi et al. (2011a). However, for higher values of AGB carbon stocks and mean AGB values, estimates lower than Saatchi et al. (2011a) were obtained. Higher estimates were obtained from this study compared to Baccini et al. (2008), with a few exceptions.

5. Conclusions

Guinea-Bissau (West Africa) is one of the poorest countries in the world, and recent developments of international conventions (e.g., UNFCCC) suggests that funding opportunities can be met through forest-based projects aimed at reducing carbon emissions through a better land use/land cover management. Therefore, an accurate and spatially explicit AGB map of the forest cover of Guinea-Bissau is paramount if carbon stocks and respective changes over time are to be quantified and assessed. This study aimed at understanding, modeling, and producing an AGB map from a relationship between AGB estimated from field data and ALOS PALSAR FBD data acquired over Guinea-Bissau (West Africa). The two proposed methods were assessed and the one derived from a BagSGB model performed better, resulting in a 10-fold cross-validated RMSE of 26.62 Mg ha^{-1}, and a correlation between observed and cross-validation predicted AGB values of 0.95. Furthermore, a map of prediction uncertainty was produced, with values of AGB coefficient of variation in the range of 7–250% (mean = 42%). Although, the results are promising, some issues have been identified and addressed.

The forests of Guinea-Bissau are inherently heterogeneous, with a high biodiversity even at the plot level. Trees with very different DBH and height in the same plot are common (e.g., Fig. 2). Also, palm trees, with large vertical trunks composed mainly of water and almost horizontal large leaves on top, are very common and dominant in the region, which adds to a more complex structure of the vegetation of Guinea-Bissau. Some limitations have been observed on the relationships between AGB and L-band SAR data in these types of ecosystems (e.g., Imhoff, 1995; Kasischke et al., 1997). The CarboVeg-GB project used a nested plot approach to estimate the AGB per forest type. Each plot had a maximum 20 meter radius, but with concentric nested plots with 4 and 14 m radius. Not all trees inside the 20 meter radius were measured and the relationship with ALOS PALSAR FBD data might have been influenced by the selected sampling design. An AGB saturation value of approximately 150 Mg ha^{-1} was observed when the BagSGB model was applied, although a lower saturation value (100 Mg ha^{-1}) was evident in Fig. 4, thus suggesting that a more complex algorithm might extract more information from the same dataset. Similarly low levels of saturation have been reported in the literature and have long been considered a limitation of L-band SAR data for the estimation of high-density values of AGB. The tested machine learning method (BagSGB) performed better than the semi-empirical model shown in Eq. (15) (adapted from Saatchi et al., 2007). These ensemble methods have already proved to be useful for the estimation of biophysical parameters (e.g., De'ath, 2007; Moisen et al., 2006). Furthermore, the BagSGB model showed better performance than a unique SGB model, and that was accomplished through a variance reduction effect. The map at the country scale obtained in this study is a first approach for delivering Guinea-Bissau with the first spatially explicit map of the distribution of AGB. The associated uncertainty at the plot level is relatively high (RMSE = 26.62 Mg ha^{-1}), but the production of a map of prediction variability (coefficient of variation) helps to identify areas of greatest uncertainty. More importantly, an uncertainty measure is placed on each pixel allowing for the estimation of uncertainty at various spatial scales. Comparison with published pan-tropical (Saatchi et al., 2011a) and continental-scale (Baccini et al., 2008) AGB maps indicates that the results presented here are more closely related to those from Saatchi et al. (2011a).

In May 2011, the ALOS completed its operations and therefore there is currently no orbital L-band SAR sensor operating. ALOS-2, which will also be carrying an L-band SAR sensor, is schedule to launch in 2013. Furthermore, NASA has practically canceled the Deformation, Ecosystem Structure and Dynamics of Ice (DESDynI) mission, which intending to put in orbit a Light Detection And Ranging (LiDAR) instrument coupled with an interferometric L-band SAR (Goetz, 2011). The BIOMASS mission, a P-band orbital SAR sensor, was proposed to the European Space Agency and selected for Feasibility Study (Phase A) in March 2009 (Le Toan et al., 2011). These and other missions are paving the way for current and future developments aimed at producing estimates of forest AGB with less uncertainty.

Acknowledgments

The *Secretaria de Estado para o Ambiente e Desenvolvimento Durável* of Guinea-Bissau and the Ministry of the Environment of Portugal funded and logistically supported the development of the Carboveg-

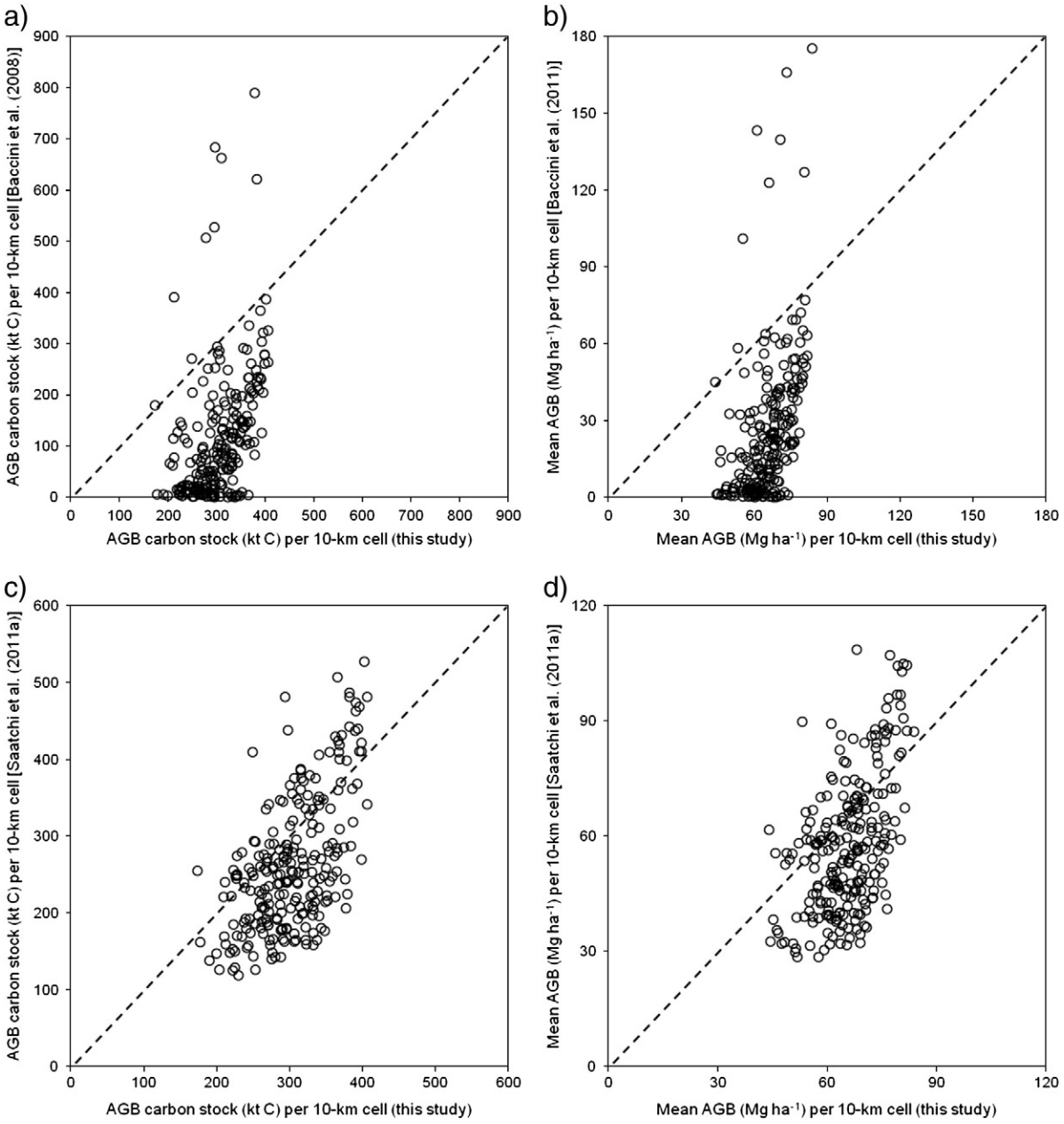

Fig. 11. Comparison of AGB carbon stock and mean AGB per 10-km cell between this study and the work of Baccini et al. (2008) and Saatchi et al. (2011a); a) AGB carbon stock and b) mean AGB, this study vs. Baccini et al. (2008); c) AGB carbon stock, and d) mean AGB, this study vs. Saatchi et al. (2011a). The dashed line represents what would be a perfect agreement relationship.

GB project. We thank all the institutions in Guinea-Bissau and the very many Bissau-Guineans who collaborated with crucial information, material, logistics, and labor during the country-wide forest data collection, especially the *Instituto para a Biodiversidade e Áreas Protegidas*, the NGO *Acção para o Desenvolvimento*, Viriato Cassamá, António Pansau N'Dafá, and Eliseu Benante. This research was conducted under the agreement of JAXA's Kyoto and Carbon (K&C) Initiative. JAXA is particularly thanked for their provision of the ALOS PALSAR data. The United States Geological Survey (USGS) is acknowledged for the provision of the Landsat 5 TM data. The GLDAS data used in this study were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC). The University of Maryland, through its Global Land Cover Facility (http://www.landcover.org) is acknowledged for their provision of the SRTM data. S.S. Saatchi (NASA Jet Propulsion Laboratory) and A. Baccini (Woods Hole Research Center) are particularly thanked for providing the biomass maps that allowed comparison with this study. The careful and insightful comments of an anonymous reviewer are particularly thanked. We highly appreciated useful comments on the various draft versions of the manuscript by Joana B. Melo (IICT).

References

Austin, J., Mackey, B., & Van Niel, K. (2003). Estimating forest biomass using satellite radar: An exploratory study in a temperate Australian eucalyptus forest. *Forest Ecology and Management*, *176*(1), 575–583.

Baccini, A., Laporte, N., Goetz, S. J., Sun, M., & Dong, H. (2008). A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environmental Research Letters*, *3*, 045011, doi:10.1088/1748-9326/3/4/045011.

Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, *36*(1–2), 105–139.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Breiman, L. (1998). Arcing classifiers. *The Annals of Statistics*, *26*(3), 801–824.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees.* Boca Raton, FL, USA: CRC Press.

Campbell, B. (2009). Beyond Copenhagen: REDD plus, agriculture, adaptation strategies and poverty. *Global Environmental Change-Human and Policy Dimensions*, *19*, 397–399.

Canadell, J., Ciais, P., Cox, P., & Heimann, M. (2004). Quantifying, understanding and managing the carbon cycle in the next decades. *Climatic Change*, *67*(2–3), 147–160.

Carreiras, J., Pereira, J., Campagnolo, M., & Shimabukuro, Y. (2006). Assessing the extent of agriculture/pasture and secondary succession forest in the Brazilian Legal Amazon using SPOT VEGETATION data. *Remote Sensing of Environment*, *101*(3), 283–298.

Carreiras, J.M.B., Lucas, R.M., & Armston, J. (Unpublished results). Assessing the effect of ALOS PALSAR incidence angle and plot size for biomass retrieval in wooded savannas, Queensland, Australia.

Castel, T., Beaudoin, A., Stach, N., Stussi, N., Le Toan, T., & Durand, P. (2001). Sensitivity of space-borne SAR data to forest parameters over sloping terrain. Theory and experiment. *International Journal of Remote Sensing*, *22*(12), 2351–2376.

Chave, J., Andalo, C., Brown, S., Cairns, M., Chambers, J., Eamus, D., et al. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, *145*(1), 87–99.

Chavez, P. (1992). Comparison of spatial variability in visible and near-infrared spectral images. *Photogrammetric Engineering and Remote Sensing*, *58*(7), 957–964.

De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology*, *88*(1), 243–251.

Dobson, M., Ulaby, F., Le Toan, T., Beaudoin, A., Kasischke, E., & Christensen, N. (1992). Dependence of radar backscatter on coniferous forest biomass. *IEEE Transactions on Geoscience and Remote Sensing*, *30*(2), 412–415.

Dobson, M., Ulaby, F., Pierce, I., Sharik, T., Bergen, K., Kellndorfer, J., et al. (1995). Estimation of forest biophysical characteristics in Northern Michigan with SIR-C/X-SAR. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(4), 877–895.

Durden, S. L., van Zyl, J. J., & Zebker, H. A. (1989). Modeling and observation of the radar polarization signature of forested areas. *IEEE Transactions on Geoscience and Remote Sensing*, *27*(3), 290–301.

Elith, J., Leathwick, J., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813.

Falkowski, P., Scholes, R., Boyle, E., Canadell, J., Canfield, D., Elser, J., et al. (2000). The global carbon cycle: A test of our knowledge of earth as a system. *Science*, *290*(5490), 291–296.

FAO (2010). *Global Forest Resources Assessment 2010. Main report.* FAO Forestry Paper 163.

FAO, UNDP, & UNEP (2008). *UN collaborative programme on reducing emissions from deforestation and forest degradation in developing countries (UN-REDD).* Framework Document.

Fransson, J., & Israelsson, H. (1999). Estimation of stem volume in boreal forests using ERS-1 C- and JERS-1 L-band SAR data. *International Journal of Remote Sensing*, *20*(1), 123–137.

Friedl, M., & Brodley, C. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, *61*(3), 399–409.

Friedl, M., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., et al. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, *114*(1), 168–182.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Friedman, J. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378.

Fromard, F., Puig, H., Mougin, E., Marty, G., Betoulle, J., & Cadamuro, L. (1998). Structure, above-ground biomass and dynamics of mangrove ecosystems: New data from French Guiana. *Oecologia*, *115*(1–2), 39–53.

Ganjisaffar, Y., Caruana, R., & Lopes, C. V. (2011). Bagging gradient-boosted trees for high precision, low variance ranking models. *34th Annual Association for Computer Machinery (ACM) Special Interest Group of Information Retrieval (SIGIR) Conference, Beijing, China*.

Gehring, C., Zelarayán, M. L. C., Almeida, R. B., & Moraes, F. H. R. (2011). Allometry of the babassu palm growing on a slash-and-burn agroecosystem of the eastern periphery of Amazonia. *Acta Amazonica*, *41*(1), 127–134.

Goetz, S. (2011). The lost promise of DESDynI. *Remote Sensing of Environment*, *115*(11), 2751.

Hansen, M., Dubayah, R., & DeFries, R. (1996). Classification trees: An alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, *17*(5), 1075–1081.

Hastie, T. R., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: data mining, inference, and prediction.* New York, USA: Springer-Verlag.

Heimann, M., & Reichstein, M. (2008). Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature*, *451*, 289–292.

Hoekman, D., & Quinones, M. (2000). Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Colombian Amazon. *IEEE Transactions on Geoscience and Remote Sensing*, *38*(2), 685–696.

Houghton, R. (2005). Aboveground forest biomass and the global carbon balance. *Global Change Biology*, *11*(6), 945–958.

Imbert, D., & Rollet, B. (1989). Phytomasse aérienne et production primaire dans la mangrove du Grand Cul-de-Sac Marin(Guadeloupe, Antilles françaises). *Bulletin d'Écologie*, *20*, 27–39.

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

IPCC (2007). *Climate Change 2007: Mitigation of Climate Change.* Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom: Cambridge University Press.

Jalabert, S., Martin, M., Renaud, J., Boulonne, L., Jolivet, C., Montanarella, L., et al. (2010). Estimating forest soil bulk density using boosted regression modelling. *Soil Use and Management*, *26*(4), 516–528.

Kasischke, E., Melack, J., & Dobson, M. (1997). The use of imaging radars for ecological applications—A review. *Remote Sensing of Environment*, *59*(2), 141–156.

Khalid, H., Zin, Z. Z., & Anderson, J. M. (1999). Quantification of oil palm biomass and nutrient value in a mature plantation. I. Above-ground biomass. *Journal of Oil Palm Research*, *II*(1), 23–32.

Komiyama, A., Ong, J., & Poungparn, S. (2008). Allometry, biomass, and productivity of mangrove forests: A review. *Aquatic Botany*, *89*(2), 128–137.

Kumar, S., Peters-Lidard, C., Tian, Y., Houser, P., Geiger, J., Olden, S., et al. (2006). Land information system: An interoperable framework for high resolution land surface modeling. *Environmental Modelling & Software*, *21*(10), 1402–1415.

Lawrence, R., Bunn, A., Powell, S., & Zambon, M. (2004). Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment*, *90*(3), 331–336.

Le Quere, C., Raupach, M., Canadell, J., Marland, G., Bopp, L., Ciais, P., et al. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, *2*, 831–836.

Le Toan, T., Beaudoin, A., Riom, J., & Guyon, D. (1992). Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, *30*(2), 403–411.

Le Toan, T., Quegan, S., Davidson, M. W. J., Balzter, H., Paillou, P., Papathanassiou, K., et al. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, *115*(11), 2850–2860.

Loarie, S., Asner, G., & Field, C. (2009). Boosted carbon emissions from Amazon deforestation. *Geophysical Research Letters*, *36*, L14810, doi:10.1029/2009GL037526.

Lucas, R., Armston, J., Fairfax, R., Fensham, R., Accad, A., Carreiras, J., et al. (2010). An evaluation of the ALOS PALSAR L-band backscatter-above ground biomass relationship Queensland, Australia: Impacts of surface moisture condition and vegetation structure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*(4), 576–593.

Lucas, R., Cronin, N., Lee, A., Moghaddam, M., Witte, C., & Tickle, P. (2006). Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, Queensland, Australia. *Remote Sensing of Environment*, *100*(3), 407–425.

Luckman, A., Baker, J., Kuplich, T., Yanasse, C., & Frery, A. (1997). A study of the relationship between radar backscatter and regenerating tropical forest biomass for spaceborne SAR instruments. *Remote Sensing of Environment*, *60*(1), 1–13.

Mitchard, E., Saatchi, S., Woodhouse, I., Nangendo, G., Ribeiro, N., Williams, M., et al. (2009). Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. *Geophysical Research Letters*, *36*, L23401, doi:10.1029/2009GL040692.

Mitchard, E. T. A., Saatchi, S. S., Lewis, S. L., Feldpausch, T. R., Gerard, F. F., Woodhouse, I. H., et al. (2011). Comment on 'A first map of tropical Africa's above-ground biomass derived from satellite imagery'. *Environmental Research Letters*, 6(4), 049001. 6, pp. 10.1088/1748-9326/6/4/049001.

Moghaddam, M., & Saatchi, S. S. (1999). Monitoring tree moisture using an estimation algorithm applied to SAR data from BOREAS. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(2), 901–916.

Moisen, G., Freeman, E., Blackard, J., Frescino, T., Zimmermann, N., & Edwards, T. (2006). Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling*, *199*(2), 176–187.

Oliver, C., & Quegan, S. (2004). *Understanding synthetic aperture radar images.* Raleigh, NC, USA: SciTech Publishing.

Oom, D., Lourenço, P., Cabral, A. I. R., Vasconcelos, M. J. P., Catarino, L., & Cassamá, V. (2009). Quantification of deforestation rates in Guinea-Bissau—A baseline for carbon trading under REDD. *33rd International Symposium on Remote Sensing of Environment (ISRSE)* (Stresa, Italy).

Pearson, T. R. H., Brown, S. L., & Birdsey, R. A. (2007). *Measurement guidelines for the sequestration of forest carbon.* General Technical Report NRS-18. Newtown Square, Pennsylvania, USA: United States Department of Agriculture, Forest Service, Northern Research Station.

Pearson, T., Walker, S., & Brown, S. (2005). *Sourcebook for land use, land-use change and forestry projects.* Arlington, VA, USA: BioCarbon Fund & Winrock International.

R Development Core Team (2011). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Ridgeway, G. (2007). Generalized boosted models: A guide to the gbm package. http://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, *85*(3), 381–394.

Rosenqvist, A., Shimada, M., Ito, N., & Watanabe, M. (2007). ALOS PALSAR: A Pathfinder mission for global-scale monitoring of the environment. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(11), 3307–3316.

Saatchi, S., Halligan, K., Despain, D., & Crabtree, R. (2007). Estimation of forest fuel load from radar remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(6), 1726–1740.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, *108*(24), 9899–9904.

Saatchi, S., Marlier, M., Chazdon, R. L., Clark, D. B., & Russell, A. E. (2011). Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. *Remote Sensing of Environment*, *115*(11), 2836–2849.

Saatchi, S. S., & Moghaddam, M. (2000). Estimation of crown and stem water content and biomass of boreal forest using polarimetric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *38*(2), 697–709.

Safavian, S., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, *21*(3), 660–674.

Sankaran, M., Ratnam, J., & Hanan, N. (2008). Woody cover in African savannas: the role of resources, fire and herbivory. *Global Ecology and Biogeography*, *17*(2), 236–245.

Santoro, M., Eriksson, L., Askne, J., & Schmullius, C. (2006). Assessment of stand-wise stem volume retrieval in boreal forest from JERS-1 L-band SAR backscatter. *International Journal of Remote Sensing*, *27*(16), 3425–3454.

Santos, J., Lacruz, M., Araujo, L., & Keil, M. (2002). Savanna and tropical rainforest biomass estimation and spatialization using JERS-1 data. *International Journal of Remote Sensing*, *23*(7), 1217–1229.

Schimel, D., House, J., Hibbard, K., Bousquet, P., Ciais, P., Peylin, P., et al. (2001). Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature, Nature*, *414*(6860), 169–172.

Schulze, E. (2006). Biological control of the terrestrial carbon sink. *Biogeosciences*, *3*(2), 147–166.

Silva, J. M. N., Carreiras, J. M. B., Rosa, I., & Pereira, J. M. C. (2011). Greenhouse gas emissions from shifting cultivation in the tropics, including uncertainty and sensitivity analysis. *Journal of Geophysical Research*, *116*, D20304, doi:10.1029/2011JD016056.

Steel, R. G. D., & Torrie, J. H. (1980). *Principles and procedures of statistics: A biometrical approach.* New York, USA: McGraw-Hill.

Suen, Y. L., Melville, P., & Mooney, R. J. (2005). Combining bias and variance reduction techniques for regression trees. *16th European Conference on machine Learning (ECML), Porto, Portugal*.

Thenkabail, P. S., Stucky, N., Griscom, B. W., Ashton, M. S., Diels, J., van der Meer, B., et al. (2010). Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using IKONOS data. *International Journal of Remote Sensing*, *25*(23), 5447–5472.

UN (1992). *United Nations framework convention on climate change.* : United Nations.

UN (1998). *Kyoto Protocol to the United Nations framework convention on climate change.* : United Nations.

UNFCCC (2011). Report of the Conference of the Parties on its sixteenth session, held in Cancun from 29 November to 10 December 2010. Addendum. *Part Two: Action taken by the Conference of the Parties at its sixteenth session*. : UNFCCC.

USGS (2006). Shuttle Radar Topography Mission, 3 Arc Second scenes SRTM_f03_p203r051, SRTM_f03_p203r052, SRTM_f03_p204r051, SRTM_f03_p204r052, SRTM_f03_p205r051, Unfilled Finished-A. *Global Land Cover Facility, University of Maryland, College Park, Maryland, February 2000.*

Werner, C., Wegmüller, U., Strozzi, T., & Wiesmann, A. (2000). *Gamma* SAR and interferometric processing software. *Proceedings of the ERS-ENVISAT Symposium, October 16–20, Gothenburg, Sweden*.

Witten, I. H., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques.* Burlington, MA, USA: Morgan Kaufmann.

Woodhouse, I. H. (2005). *Introduction to microwave remote sensing.* Boca Raton, FL, USA: CRC Press.