Assessing Forest Biomass and Exploration in the Brazilian Amazon with Airborne InSAR: an Alternative for REDD

Julianno B. M. Sambatti*, Raphael Leduc, Dieter Lübeck, João Roberto Moreira, and João Roberto dos Santos

Timberline Ltda, São José dos Campos Area, Brazil

Abstract: REDD (Reduced Emissions from Deforestation and forest Degradation) constitutes a set of financial incentives designed to reduce CO_2 emissions from forest degradation and deforestation. REDD success depends on measuring forest biomass as a proxy for CO_2 stocks. We tested the efficacy of airborne X- and P-band interferometry as a remote-sensing method to quantify forest biomass and detect changes in forest structure in the Paragominas region, eastern Amazon. With field-classified regions of interest (ROI) and radar imagery, we classified an area of 1479.66 km2 into four forest classes. Radar backscatter and interferometric variables of each forest class were statistically examined. We obtained the interferometric height, H_{int}, by subtracting digital elevation models resulting from X and P band interferometry for the study area. Inventory-measured biomass were obtained for 42 plots nested within these forest classes, and used as ground truth for subsequent analyses. Using these field plots as experimental units, a functional relationship between radar variables and above-ground biomass (*AGB*) was obtained by fitting a linear model relating inventory-measured *AGB* and radar-derived variables. A map of *AGB* was created. Combining backscatter variables and H_{int} effectively classified the forest. *AGB* can be predicted for the Paragominas landscape with H_{int} and the P-band polarizations *pHV* and *pVV* (R^2=0.82, normalized RMSE=13.7%). X- and P-band airborne radars can be used to estimate forest *AGB* of large continuous areas as well as detecting forest exploration and degradation, events REDD intends to prevent.

Keywords: REDD, SAR, forest biomass, forest carbon stock, above ground biomass, RADAR, classification.

INTRODUCTION

REDD (Reduction of Emissions from Deforestation and Degradation) is a mechanism designed to prevent CO_2 emissions by which legal holders of forested areas are paid to avoid forest degradation and deforestation (REDD; http://unfccc.int/methods_science/redd/items/4531.php, http://www.un-redd.org/). REDD has become a critical issue in international discussions about post-Kyoto protocol compliance mechanisms, and is becoming very relevant in tropical countries where forests abound [1], and where deforestation causes substantial CO_2 emissions [2]. Similar to the CDM (Clean Development Mechanism, http://cdm.unfccc.int/), the concept of REDD has a focus on projects whose activities are expected to reduce CO_2 emissions within project limits given a deforestation baseline. In addition, leakage [3, 4] – *i.e.*, deforestation off project limits resulting from REDD project activities – is a major concern that project proponents need to address before a project is approved as a REDD project. Guaranteeing that projects will reduce emissions and leakage will not take place are conditions for REDD to be successful. Thus, employing accurate, cost-effective, and robust wall-to-wall methods – *i.e.*, methods capable of surveying an entire target area, and not relying on samples taken to represent that area – to quantify and detect changes in forest carbon stocks has

become quite important to attend demands from REDD implementation [5-7]. Since forests store CO_2 in their biomass, measuring the forest biomass is the most straightforward method to quantify changes in carbon stocks in forests, and several methods to measure forest biomass are either available or under development [5, 6].

Performing forest inventories in sample plots within the project area is a consolidated method of forest biomass estimation [8-11]. However, its extremely complicated logistics and high costs prevent the widespread use of forest inventories for the wall-to-wall monitoring of carbon stocks in large and continuous areas. Forest biomass (carbon stocks) can also be evaluated using remote sensing instruments on airborne or satellite platforms, but substantial refinements are needed before routine assessments can be made at national or regional scales [12].

Satellite imagery has been used to assess historical deforestation rates in the Brazilian Amazonia through satellite images provided by INPE (the Brazilian National Institute for Space Research) using optical data (projects Prodes and Deter [13]), which is an analysis somewhat limited by cloud cover in the tropics [14]. This is a qualitative analysis that only indicates the presence or absence of forests in the landscape. More quantitative analyses to estimate forest carbon stocks indirectly by developing statistical relationships between ground-based measurements and satellite-observed vegetation indexes have been attempted [15]. However, the current suite of optical satellite sensors, such as Landsat, AVHRR and MODIS,

*Address correspondence to this author at the Timberline Ltda, São José dos Campos Area, Brazil; Tel:+55 12 9603 2272; E-mail: jbmsamba@gmail.com
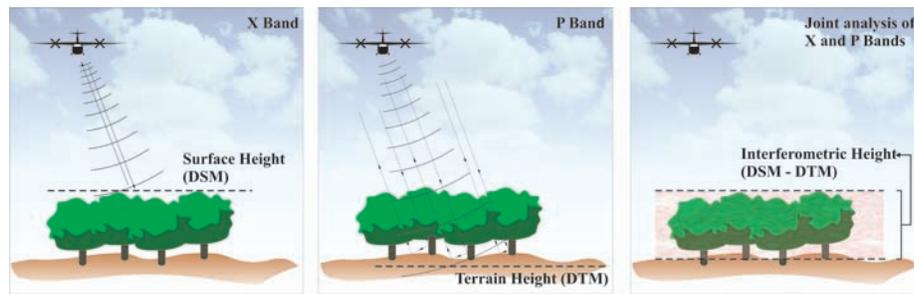
Fig. (1). Schematic view of the airborne radar data acquisition. P-band and the DTM, X-band and the DSM, and DSM-DTM obtaining the interferometric height, H_{int}.
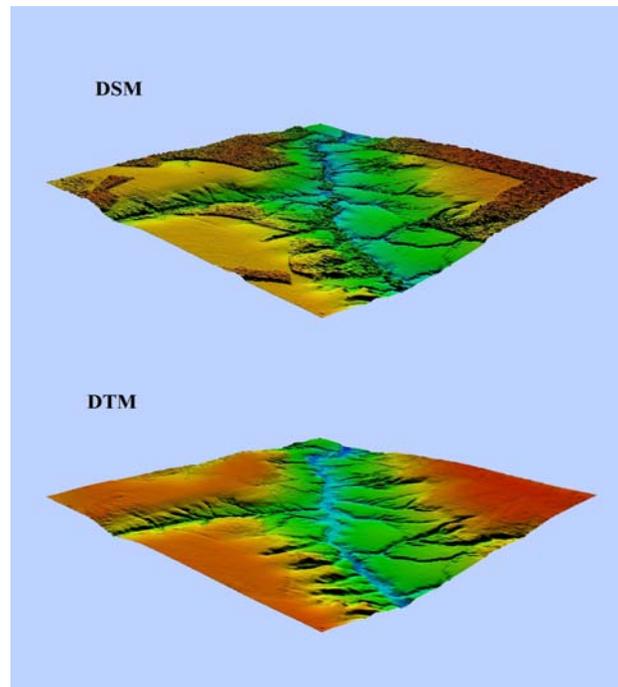
Fig. (2). Example of a DSM (with the landscape partially covered with forests) and a DTM (bare landscape) obtained in the region of Paragominas.

cannot yet be used to estimate carbon stocks of tropical forests with certainty [16].

LiDAR is a remote sensing technology able to measure the distance to a target by illuminating this target with pulses of laser. LiDAR can use the laser beam reflection time from a source (such as an airplane) to a target (such as a tree) to create a tri-dimensional model of a forest. Although, LiDAR remote sensing can potentially exceed the capabilities of radar and optical sensors to accurately estimate carbon stocks for all forest types [14, 17-21], LiDAR presents some limitations that may restrict its broad application. Cloud cover or the presence of dense canopies, for example, also limit the gathering of LiDAR data, which can become a problem in tropical rainforest regions.

Radars constitute a remote sensing alternative that can provide topological information about forest features through the analysis of microwave backscattering even in the presence of cloud cover. Quantifying carbon stock-related forest features in relatively homogeneous or young forests has been attempted with synthetic aperture radar (SAR) sensors on board of several satellites (*e.g.*, ALOS/PALSAR [22], Envisat [23], RADARSAT-2 [24], TerraSAR [25]). However, the radar signal tends to saturate when forest

biomass exceeds 50 – 300 Ton/ha, depending on study conditions [26-34], which are lower values than those often observed in mature Amazon forests [35-39]. Because different radar bands (*e.g.* X-, C-, L-, P-bands) provide different information about forest features, they can be wisely combined to overcome such limitations. The combination of different radar bands to perform the interferometric analysis of forest biomass will be optimal if it provides information of biological relevance, such as allometry-related parameters (*e.g.*, forest height). With that in mind, X- and P-bands offer several advantages for forest interferometry. Airborne X- and P-band interferometry has been extensively employed in Brazil for mapping purposes (http://www.dsg.eb.mil.br/index.php/projetos-e-convenios/em-andamento/radiografia-da-amazonia?start=2), and some attempts have been conducted to use such systems to estimate forest biomass [40, 41]. As opposed to L- and C-bands that penetrate forests in different degrees, which makes the interpretation of interferometric data problematic, X-band waves reflect from the vegetation canopy or bare ground and can be used to generate Digital Surface Models (DSM) through interferometry (Figs. **1** and **2**), and P-band waves can fully penetrate the forest canopy, reflecting from the forest floor and tree trunks by the effect of double

Table 1. Acronyms and Symbols That Appear in the Text

Emission Reduction Mechanisms	
REDD	Reduced Emissions from Deforestation and forest Degradation
CDM	UN Clean Development Mechanism
Forestry	
AGB	Above Ground Biomass
ρ	Wood Density
DBH	Diameter at Breast Height
H	Tree Height
h_{dom}	Dominant Height
h_{sub}	Submergent Height
Remote Sensing	
SAR	Synthetic Aperture Radar
InSAR	Interferometric Synthetic Aperture Radar
LiDAR	Light Detection And Ranging
DSM	Digital Surface Model
DTM	Digital Terrain Model
H_{int}	Interferometric Height
pHH	P-band Polarization, Horizontal – Horizontal
pHV	P-band Polarization, Horizontal – Vertical
pVV	P-band Polarization, Vertical – Vertical
xHH	X-band Polarization, Horizontal – Horizontal
CHV	C-band Polarization, Horizontal - Vertical
LHH	L-band Polarization, Horizontal - Horizontal
LHV	L-band Polarization, Horizontal – Vertical
LVV	L-band Polarization, Vertical – Vertical
Classification	
FD	Dense Forest
FE	Explored Dense Forest
FS	Secondary Forest
SS	Secondary Succession
ROI	Region of Interest
J-M distance	Jeffries-Matusita Distance
Statistics	
RMSE	Root Mean Square Error
R^2	Coefficient of Determination
SD	Standard Deviation
AIC	Alkaike Information Criterion
BIC	Bayesian Information Criterion
ε	Linear Model Error
α and β_n	Linear Model Parameters

bounce (Fig. **3**). This is an effect that depends on the presence of vegetation to occur. Without vegetation, P-band waves reflect away from the radar source and no radar signal can be recorded [42]. In the presence of vegetation, P-band waves reflect back to the radar source by encountering, for example, a tree trunk barrier, and producing radar backscatter. P-band interferometry is used to determine the Digital Terrain Model (DTM) (Figs. **1** and **2**). As a result, the forest height can be obtained by analyzing the difference between DSM and DTM, the interferometric height difference (H_{int}) (Fig. **1**).
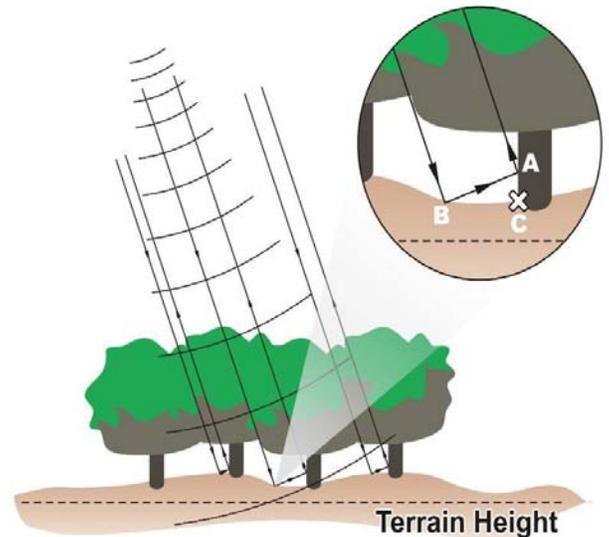

Fig. (3). Schematic view of the P-band double bounce effect. Due to double reflection A and B, the real reflection point is C, at the base of the trunk, which represents terrain altitude

Since the variation in P-band backscatter intensity constitutes additional information about forest structure and biomass, the joint analysis of forest canopy height data obtained with radar interferometry together with backscatter-related variables has been shown to substantially improve the prediction of forest biomass and forest structure from radar data (see [31, 40, 41, 43-45]). For example, areas cleared recently should show a shorter interferometric height than mature and dense rainforests due to their shorter forest canopies. In another instance, one expects that an explored forest will present a smaller P-band backscatter than a non-explored forest due to the removal of large trees because, without the barrier formed by tree trunks in an explored forest, a greater proportion of P-band waves will reflect away from the radar source [42].

Here we evaluate the performance of X- and P-band airborne InSAR – Interferometric Synthetic Aperture Radar – with lateral view (Fig. **1**) in estimating forest biomass and detecting changes in forest structure in forests of eastern Amazon. In this study, we explored a data set obtained in the municipality of Paragominas, state of Pará, Brazil. This data set comprises field classified samples, field inventories, and radar data for the entire study region. The objectives were: 1) To study the individual contributions of radar X- and P-band variables for forest classification. 2) Determine which tree-height stratum best predicts H_{int} in Paragominas forests. And, 3) statistically define an optimal functional relationship between radar-derived variables and field inventory-

Table 2. Flight Parameters for Radar Data Acquisition

Flight Parameter	Value
Baseline for X-band interferometry equals to the distance between the two antennas	2.77 m
Baseline for P-band interferometry	40 m
Flight altitude	12000 feet
East-West flight track	7 km
North-South flight track	2 km
Flight overlap	66.00%
Resolution of the DSM	1.0 m x 1.0 m
DSM pixel size	0.5 m x 0.5 m
Resolution of the ORI X-band	0.5 m x 0.5 m
ORI X-band pixel size	0.5 m x 0.5 m
Resolution of the DTM: 1.0 m x 1.0 m w/o vegetation, 5.0 m x 5.0 m with vegetation	1.0 m x 1.0 m w/o vegetation, 5.0 m x 5.0 m with vegetation
DTM pixel size	2.5 m x 2.5 m
Resolution of the ORI P-band: 2.5 m x 2.5 m	2.5 m x 2.5 m
ORI P-band pixel size	2.5 m x 2.5 m

estimated forest biomass (used as ground truth) for eastern Amazon. To help interpretation, Table **1** shows acronyms and symbols appearing in the text.

MATERIALS AND METHODOLOGY

Study Area

This study was conducted in an area of 1479.66 km^2 in the Paragominas region, state of Pará, Brazil, between 2° 45' S and 3° 00' S, and 47° 15' W and 47° 30' W (Fig. **4**). The climate in this region is equatorial humid with a drier season between June and November, and a rainy season between December and May. Annual temperatures average 26° C, relative humidity 81%, and annual precipitation is approximately 1800 mm [46]. The region's original vegetation is typically Amazonian lowland rainforest. Today, the Paragominas landscape consists in a mosaic of this original forest interspersed with patches of explored forests in variable degrees, agricultural land, and pastures.

Airborne Radar Data Acquisition

The radar OrbiSAR-1 (OrbiSat) was installed in a Rockwell Turbo Commander 690 B aircraft. By flying overlapping tracks above the study area, this system is capable of acquiring complex images from X- and P-bands, including phase difference values related to the terrain and surface. For this study, we used X-band single pass interferometry with three antennas installed in the aircraft, and a multipol single antenna for P-band that requires repeat pass to perform P-band interferometry. Flight parameters can be seen in Table **2**.

Fig. (4). Map of Brazil with the municipality of Paragominas within the Legal Amazon.

After the processing of interferometric radar raw data, an X-band mosaic, including X-band orthoimages and Digital Surface Model (DSM), and a P-band mosaic, including multipolarimetric P-band orthoimages and Digital Terrain Model (DTM) were generated. For each image pixel, the difference between the values of DTM and DSM represents the vegetation interferometric height (H_{int}) [40, 41].

Supervised Landscape Classification with Airborne SAR Imagery

[47] Provides a general guideline for landscape classification via remote sensing techniques used in this work. For the present study, a random sample of 630 points for ground truth was taken in the 1479.66 km^2 of the study area in the Paragominas region. This sampling aimed at representing all land use/land cover types found in the entire study area. The field crew classified the vegetation around each sampled point, and indicated ROIs (Regions of Interest) where the surrounding area was large enough to represent the classified vegetation in the available radar images. The coordinates of these ROIs were recorded. Flights to collect radar data were performed between April 23rd and May 1st of 2005. Multipolarimetric P-band orthoimages, X-band orthoimages, coherence images, and digital terrain and surface models of the landscape (Digital Terrain Model from P-band, DTM and Digital Surface Model from X-band, DSM) for the entire study area were generated as described above.

An initial classification considered each ROI as a class. A distance matrix using the Jeffries-Matusita (J-M) distance was constructed and used to condense ROI classes into classes with the same quantitative behavior. The J-M distance varies from 0 to 2, measures the degree of dissimilarity between classes [48, 49], and is widely used in the quantitative classification of the vegetation of heterogeneous landscapes using remote sensing tools (*e.g.*, [50, 51]). The mean of the J-M distance within a class, and with respect to all other classes were taken as a reference to

decide whether or not ROIs should be eliminated from the analysis. Two ROIs were considered suspicious and subject to a more careful scrutiny when the J-M distance between them was consistently too high with respect to the mean of the class they were assigned to. Some ROIs were eliminated from the analysis if the H_{int} was clearly not consistent with the expectations. For example, one of the eliminated exposed soil ROIs showed a large H_{int} (expected to be zero), and its J-M distance to other ROIs of that class was consistently higher than the mean distance between other pairs of exposed soil ROIs. Pairs of ROIs were considered to be unequivocally separable when J-M>1.4 (see [47]). After applying this algorithm, ROI classes were condensed into only eight land use/land cover classes. Non-forest classes included: exposed soil, abandoned (dirty or invaded by non-grass species) pastures, clean pastures, water bodies or wet areas. Forests were classified into four broad classes: 1) Dense forest (FD), with the dominance of large individuals representing the original vegetation. 2) Explored Dense Forest (FE), a forest type resulting from a recent selective harvesting. Here, the label Explored Forest (FE) does not mean sustainable explored forest; a visual assessment of this vegetation class suggests that these FEs were rather heavily explored. 3) Secondary Forest (FS), resulting from the natural forest regeneration after clearing for shift-cultivation and/or pastures. And, 4) Secondary Succession (SS), a forest in early/intermediate successional stages resulting from the replacement of natural vegetation re-growth, and where grasses, ferns, and pioneer species are dominant.

To quantify the classification accuracy, a confusion matrix using a subset of samples of these eight classes used for classification was constructed and the Kappa coefficient, which varies from 0-1 and measures the degree of coincidence when the same object (sample) is assessed in two different ways (classification algorithm and ground truth), was calculated.

With these pre-defined classes, the classification training was performed with an image classification algorithm. The trained algorithm was used for a supervised land use/land cover classification of the entire landscape based on how the airborne InSAR imagery was associated with the field classified samples using the maximum likelihood classifier [52-54]. The maximum likelihood classification assumes that the statistics for each class are normally distributed and uses this property to calculate the probability that a given pixel belongs to a specific class. A pixel is assigned to the highest probability class (ENVI Tutorial: Classification). These analyses were performed with the software ENVI (ITT).

Field Inventories

An imaginary circle with a diameter of 24 km encompassing all four forest classes limited the area sampled to allocate forest inventories; *i.e.*, approximately 452 km^2. Forest inventories were conducted between November and December of 2008.

Inventories were performed in a total of 8.6 ha divided into 42 plots randomly assigned and nested within these four different forest classes defined based on the classification described above: FD, 12 plots; FE, 12 plots; FS, 8 plots; and SS, 10 plots. The dendrometric parameters of FE and FD were obtained from the conglomerates with 4 rectangles of

2500 m^2 (20m x 125m) in each plot. Plots of this size arranged in conglomerates are known to capture the environmental variation satisfactorily [55]. FS plots were rectangles of 2000 m^2 (20 x 100 m). SS plots were rectangles of 1000 m^2 (20 x 50m). In contrast to FD and FE (using conglomerates), all FS and SS plots were randomly assigned within their respective forest class regions. Plot size was proportional to the degree of structural feature variation in the vegetation. For example, the number of new species stabilized after sampling five SS plots (data not shown). By designing smaller plots for more homogeneous vegetation, we were able to increase the number of plots of this forest class. Sample plots with these areas (and even smaller) have been used in field inventories to study forest attributes (including biomass) in the tropics [41, 56-59]. Importantly, note that we did not use these plots to infer the biomass of a given region. These plots were used here to establish a statistical relationship between radar variables and field-estimated biomass.

We measured *DBH* (tree diameter at 130 cm above forest floor) for each tree within these plots, but recorded only the diameter of trees with *DBH* >10 cm in FD, FE, and FS plots, and *DBH* > 5 cm in SS plots. A laser distance meter device (Bosch DLE-50) was employed to measure total tree height for every tree whose *DBH* was recorded. There were difficulties to visualize the laser beam on the top of some trees. Trees that could be measured unequivocally were used to help reducing uncertainties when measuring those more difficult trees [60].

GPS coordinates were difficult to obtain in closed forests with the available technology at the time of data acquisition. A position within each plot was identified where coordinates were possible to be obtained with a GPS Garmin (Etrex). From these coordinates, the rectangle corners were determined geometrically. Plot corner coordinates were determined with ArcMap. A shape file was created.

With a shape file containing the coordinates of each field-inventoried plot, we obtained the statistics (*i.e.*, mean and SD) for each radar image (backscatter) and the H_{int} per plot. Each plot was characterized by the across-pixel average of H_{int} in meters, and of the P-band backscatter values for the polarizations – *i.e.*, the orientation of the electric and magnetic fields of the microwave – *pHH*, *pVV*, and *pHV*, as well as the X-band backscatter value *xHH*, all of them expressed in decibels.

Exploratory Data Analysis on Radar Variable Contribution to Landscape Classification

Since forest classification used field-classified samples to train a maximum likelihood classification algorithm to classify the entire landscape based solely on radar data, which radar variables or combination of variables were employed by this algorithm to perform such classification are unknown *a priori*. To understand the role played by the different radar variables in the classification process, and their ability to detect differences in forest structure, we created boxplots for each radar variable per forest class – *i.e.*, FD, FE, FS, or SS – including per plot mean interferometric height (H_{int}), *pHH*, *pVV*, *pHV*, and *xHH*. Note that, although plot area varied from class to class, the mean of all pixels in a plot covered with a particular forest class is a quantity

independent of its area. Differences in plot area will only affect the number of pixels employed to obtain the mean.

Allometric Corrections for Estimation of Aboveground Forest Biomass

Because the flight to obtain radar data took place three years before the forest inventory, the inventory data had to be corrected to account for forest growth along these three years.

This correction was performed using an empirical approach described in [61], who developed equations to estimate secondary forest growth. Since FD is an older forest, it was assumed that FD growth was negligible along these three years, and corrections were only applied to the SS, FS, and FE classes. Whether or not correcting the explored forest biomass FE was subject to some uncertainties; the intensity of forest exploration, for example, is unknown. This information about forest exploration has a large impact on whether the remaining forest is closer to a dense forest (FD) or, alternatively, to a secondary forest (FS). Thus, a preliminary analysis was performed with and without correcting the explored forest (FE) biomass. After comparing the boxplot of interferometric height by forest class with the boxplot of Above Ground Biomass by forest class (see Results section), we decided that the explored forest should be treated as a secondary forest, and the statistical analysis was carried out with the FE data corrected for growth during the three years elapsed between flight and forest inventory.

The correction for forest growth between flight and field inventories was applied as follows: 1) Using the equation (6) in [61], tree age was estimated using the tree height. 2) With equations (9, 12, 17 and 6) in [61], each tree height and *DBH* were re-estimated using the estimated age minus three years (the gap between flights and field survey). 3) Each tree was taxonomically identified and, using a public data base [62, 63], its wood density ρ in g/cm^3 was obtained. Including wood density can significantly improve the allometric estimation of the forest biomass [10, 11]. When the species was not listed in the data base, the average wood density of other species in the genus available in the data base was used. For the three species without any register at the genus level, the average wood density of all species (0.558) was used. 4) Tree above-ground biomass was estimated as in [11] using the formula for moist forest:

Tree Above-Ground Biomass *AGB* (Ton) = EXP[-2.977 + ln ($\rho * DBH^{2} {}^{*} H$)] [11]   (1)

where

H = tree height (m)

DBH = diameter breast height (cm)

and ρ = wood density (g/cm^3)

Tree biomass after correction that reached a value close to zero was considered negligible and the tree was discarded from the analysis. Biomass per plot was estimated by adding the biomass of all trees in the plot. Because plots had different areas (1000, 2000, and 2500 m^2), a standardized biomass in Tons per hectare was calculated for each plot.

Table 3. **Total Number of Polygons, Area in km^2 and % of the Classified Area Per Land Cover/Land Use Class**

Class	Polygons	Area (km^2)	%
Exposed soil	9426	90.65	6.13
Abandoned (dirty) Pasture	6109	249.17	16.84
Clean Pasture	3532	35.51	2.4
FE	20646	293.46	19.83
SS	12763	478.47	32.34
FD	17866	188.39	12.73
FS	20776	139.93	9.46
Water bodies – Wet areas	1323	4.08	0.28
Total Classified Area	92441	1479.66	100

Statistical Relationship Between Forest Above-Ground Biomass (*AGB*) and Radar Variables

Assessing the Impact of Different Tree Strata on the Interferometric Height (*H$_{int}$*)

Although H_{int} is highly correlated with canopy height, not all trees determine the canopy height. It has been demonstrated [41] that, while canopies of dense Amazon forest in the Tapajós region are predominately formed by the top 20% highest trees – the dominant height (h_{dom}) –, forests at early successional stages have canopies formed by the shortest 20% trees - the submergent height (h_{sub}). From this analysis, it has been concluded [41] that H_{int} is estimated based on a different subset of trees depending on the forest successional stage. To examine these patterns in forests of the Paragominas region, we have performed two linear regressions: 1) regression of H_{int} on the subset of trees representing the tree height upper 10% percentile, and 2) regression of H_{int} on the subset of trees representing the tree height lower 20% percentile. By finding which tree subset best predicts H_{int} – *i.e.*, better matches H_{int} – one can obtain an indication of which tree stratum most contributes to H_{int}. If, for example, the mean tree height of a tree stratum (say the shortest 20% of all trees) in a given area is 10 m, but H_{int} is 20 m in the same area, there is a clear disconnection between this tree stratum and H_{int}. If another tree stratum (say the tallest 10% of all trees) in that same area shows a mean height closer to 20 m, one can conclude that this tree stratum contributes more than the shorter stratum to forming the forest canopy and, consequently to H_{int}.

Forest Aboveground Biomass (*AGB*)

Dimensional reasoning [64] suggests that above ground biomass (*AGB*) should scale as a power of H_{int}. For a single tree, this scaling is usually accomplished through the equation

$AGB = a \, (tree \; height)^{b}$ (2)

(*e.g.*, [65, 66]), which can be linearized for analytical purposes as $\log_{10} AGB = \log(a) + b \log_{10}(tree \; height)$, and whose parameters *a* and *b* can be estimated statistically. Based on this concept, we used a similar scaling to relate H_{int} and *AGB*: $\log_{10} AGB = \log(a) + b \log_{10} H_{int}$. All subsequent statistical analyzes were conducted in this linear form.

Fitting statistical linear models was employed to determine which combination of radar attributes can best explain forest biomass. We initially tested a model with the $\log_{10}AGB$ as the response variable and $\log_{10}H_{int}$, pHH, pVV, pHV, and xHH as independent variables. Stepwise model selection algorithms (function *stepwise* of R) were applied to select the optimal linear model [67] to estimate AGB using radar variables. The normality of residuals of all models was tested with the Shapiro-Wilk normality test (function *shapiro.test* in R). And, diagnostic plots were obtained for all models to verify whether model assumptions have been attended. A k-fold (k=10) cross-validation algorithm (function *CVlm* in R) was used to verify how robust was the selected model in testing data sets. The overall mean square of the k-fold analyses (in a log scale) was estimated. The RMSE of the model using the entire data set was estimated. All statistical analyzes were carried out with the software R [68].

Map of the Forest Aboveground Biomass

An area of 197 km^2 covered with forest in the Paragominas region was selected. The per-pixel biomass was calculated with the formula derived from the statistical model obtained with the statistical analyses. The geo-referenced image was generated showing the biomass classification. The image was created with pixels representing a biomass in Ton/ha. The map includes the calculation of the biomass of a selected rectangle of 853 ha (with ENVI) by adding up the absolute biomass in Tons (corrected for the pixel area of 6.25 m^2) estimated for each pixel within the area.

RESULTS

Radar Attributes in Forest Classification

Table **3** shows the classified land use/land cover classes, the number of polygons, area and % of each class. The trained classification algorithm was able to confidently separate four forest classes: FD (Dense Forest), FE (Explored Dense Forest), FS (Secondary Forest) and SS (Secondary Succession) (Fig. **5**). The Kappa coefficient associated with the confusion matrix (APPENDIX **1**) constructed with these eight classes was 0.9026, close to one, indicating a good agreement between radar and field data.
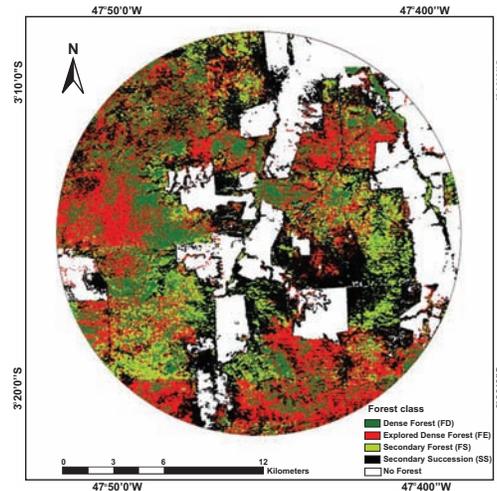

Fig. (5). Distribution of forest classes in the Paragominas landscape from classified images obtained with the Orbisar-1 sensor.
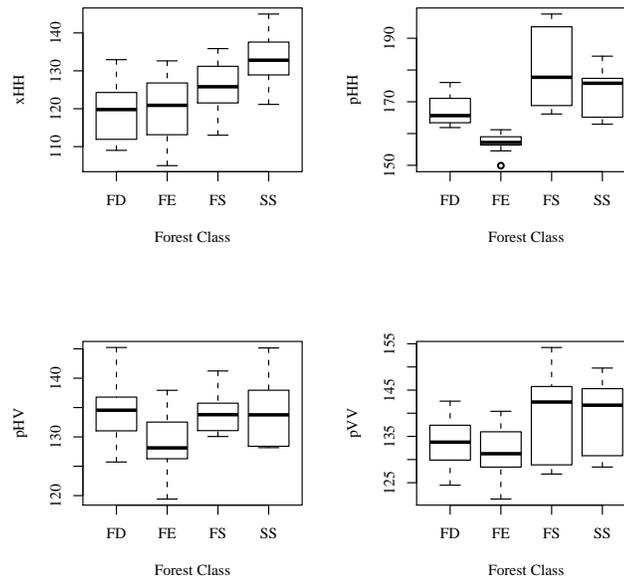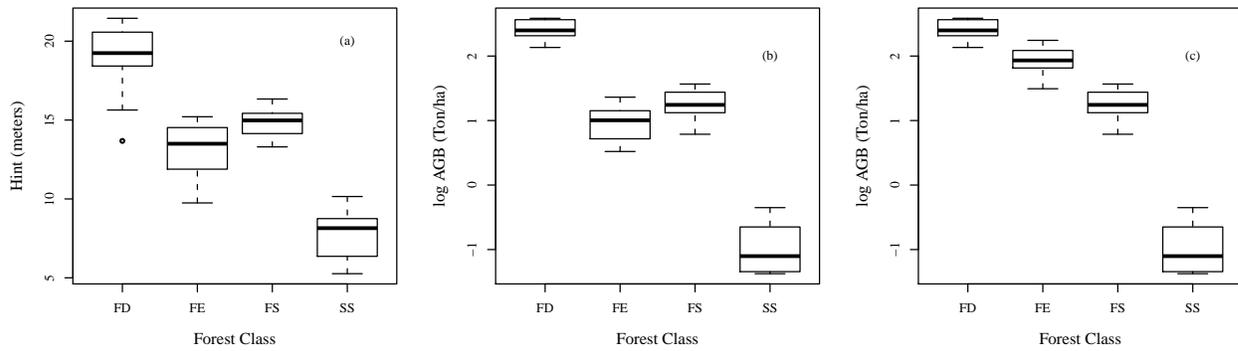

Fig. (6). Boxplots of backscatter variables - *xHH*, *pHH*, *pHV*, and *pVV* - by forest class - FD, FE, FS, and SS. Backscatter unit is in decibells and is proportional to the area of obstacles encountered by P- and X-band microwaves.

Fig. (7). Comparison between (a) boxplot of H_{int} , (b) $\log_{10}AGB$ with FE biomass corrected to account for the growth during the three year gap between flight and inventory, and (c) $\log_{10}AGB$ without such FE biomass correction by forest class.
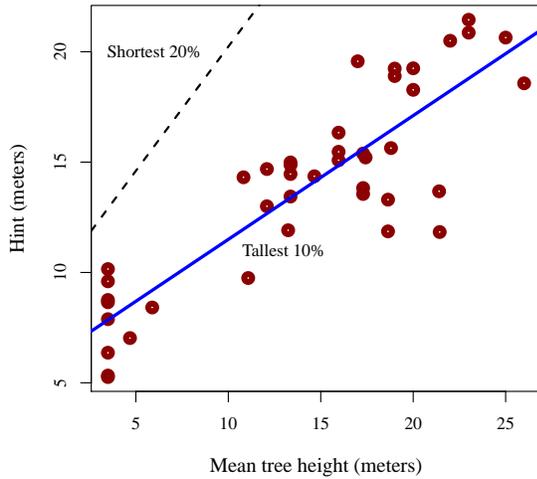
Fig. (8). Scatterplot of H_{int} by the per-plot average of the tallest 10% trees per plot with a fitted continuous line representing the regression of H_{int} on this tree subset. The broken line represents the regression of H_{int} on the per-plot average of the shortest 20% trees.
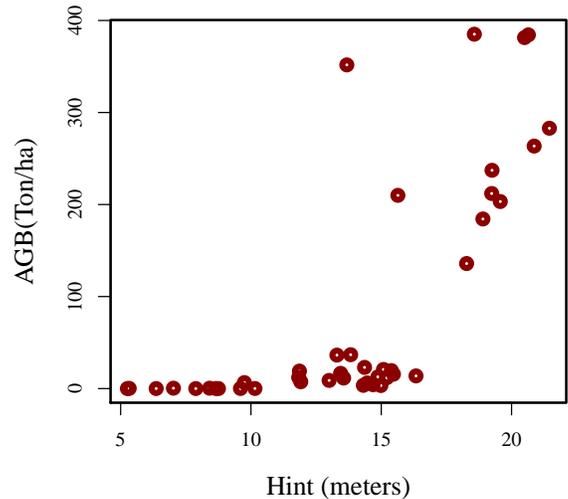
Fig. (9). Scatterplot of the forest AGB per plot in Ton/ha estimated via forest inventories and the H_{int} in meters obtained via radar interferometry using X- and P-band.

When employed by the trained classification algorithm to classify the entire landscape, radar variables contributed differently to distinguishing plots representing these forest classes. The boxplots of *pVV* and *pHV* by forest class showed that these radar variables were not different across forest classes (Fig. **6**). The variable *xHH* showed increasing values as forest classes move from FD to SS, and sets SS apart from FE and FD (Fig. **6**). The variable *pHH* was not different when one compares FS, SS and FD, but clearly shows a smaller mean and seemingly a smaller variation in FE when compared to the other classes (Fig. **6**).

Fig. (**7**) displays three boxplots: (a) H_{int} by forest class, forest biomass by forest class with (b) and without (c) FE biomass correction for forest growth between flight and field inventories, as described in the Methods section – *Allometric corrections for estimation of aboveground forest biomass*. Correcting for forest biomass growth produced a large impact on the FE biomass. Without such correction, FE biomass is intermediate between FD and FS. With the correction, FE biomass is, on average, very close to the FS biomass. By comparing these boxplots, one can notice that the biomass boxplot with corrected FE (b) is very similar to the boxplot depicting H_{int} per forest class (a); H_{int} parallels forest biomass in this case. Several studies involving forest

biomass estimation with InSAR showed that H_{int} is the most important variable in the forest biomass estimation [40, 41]. This and the visual evidence that FE has been rather heavily explored (data not shown) suggest that FE is more similar to FS than it is to FD, and that correcting FE biomass is adequate.

Interferometric Height (H_{int}) Versus Tree Height Strata

Because H_{int} reflects the height of the forest canopy, the tree stratum – whether the tallest or shortest trees (or both) – that best predicts H_{int} will show a mean height closer to H_{int}. Fig. (**8**) shows two regression lines. The regression line of H_{int} on tree height of the tallest 10% trees shows that tree height in this tree stratum matches much better H_{int} than the tree height of the shortest 20% trees (regression line represented by the broken line in Fig. **8**). Thus, in general, the tallest 10% trees better predict H_{int} than the shortest 20% trees. Note that we tested a regression of H_{int} on the tallest 20% trees (data not shown), but opted to perform the regression on the tallest 10% trees because this tree subset presented an even closer match with H_{int} values than the subset of tallest 20% trees. Nonetheless, this better prediction is not constant as we move from shorter (corresponding to SS plots) to taller (corresponding to FD

Table 4. Statistical Models with their Coefficients, Coefficient Standard Errors (SE), the p-Value Associated with these Estimates, Model Multiple and Adjusted R^2, Model Residual Standard Error (RSE), and Model p-Value

Model 1 – $\log_{10}AGB = \log(\alpha) + \beta_0 \log_{10}H_{int} + \beta_2 pVV + \beta_3 pHV$				
	Intercept (Log (α))	β_0	β_2	β_3
Coef.	-6.54	6.94	-0.03	0.03
SE (p-value)	2.26 (0.0063)	0.54 (<0.0001)	0.02 (0.0481)	0.01 (0.0057)
	Multiple R^2 = 0.83	Adjusted R^2 = 0.81	AIC = 74.95	
	RSE: 0.551 on 38 degrees of freedom	Model p-value<0.0001		

Model 2 – $\log_{10}AGB = \log(\alpha) + \beta_0 \log_{10}H_{int} + \beta_5 (pHV - pVV)$			
	Intercept (Log (α)	β_0	β_5
Coef.	-6.68	6.94	-0.03
SE (p-value)	0.60 (<0.0001)	0.53 (<0.0001)	0.01 (0.0033)
	Multiple R^2 = 0.83	Adjusted R^2 = 0.82	AIC=72.95
	RSE: 0.544 on 40 degrees of freedom	Model p-value<0.0001	

plots) canopies. Because lines diverge from model intercepts (slopes = 5.88 and 8.97) – intercepts which are not significantly different from each other (intercept standard errors overlap) –, the contribution of both smaller and taller trees to predicting H_{int} (*i.e.*, forming the forest canopy) are similar and equally important in early succession forests (SS), but the contribution of shorter tree strata to H_{int} decreases continuously as forests become taller and older (FD).

**Statistical Model Relating *AGB* and Radar Variables**

The statistical models relating *AGB* and radar variables were derived from field plots taken from several forest classes whose inventoried biomass range from 0.042 (SS) to 385.077 (FD) Ton/ha after applying the appropriate corrections described above. A scatterplot of biomass in Ton/ha per H_{int} in meters was created (Fig. **9**).

The initial statistical model was fitted using the following equation based on the linearized allometric equation described for a single tree height (equation 2) plus radar backscatter variables:

$\log_{10}AGB(Ton/ha) = \log(\alpha) + \beta_0 \log_{10}H_{int} + \beta_1 pHH + \beta_2 pVV + \beta_3 pHV + \beta_4 xHH + \varepsilon$ (3)

In this linear model, $\log(\alpha)$ is the model intercept. Had H_{int} been the only variable in the model, (α) would represent the parameter (*a*), and β_0 the parameter (*b*) of the allometric equation $AGB = a\ H_{int}^b$. ε is the error. The initial statistical model including all radar attributes was highly significant (p<0.0001) with an adjusted R^2= 0.80.

The resulting model was considered optimal to estimate forest biomass using radar parameters. After applying the model selection algorithm (using both AIC and BIC), the resulting simplified model was still highly significant (p < 0.0001) with an adjusted R^2 = 0.81. Normality of residuals
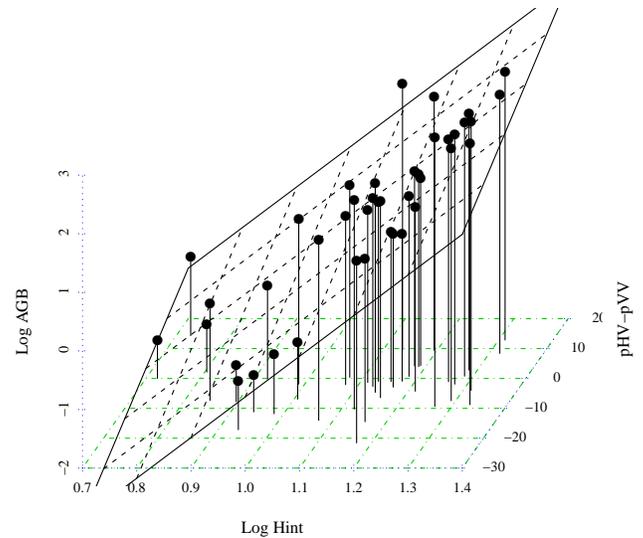
Fig. (10). 3D scatterplot of the $\log_{10}AGB$ as a function of $\log_{10}H_{int}$ and the compounded variable (pHV-pVV) with a fitted plane representing the linear model described by the equation 6.

was confirmed for both models (p=0.76 and p=0.56 respectively). Diagnostic plots were obtained to verify whether model assumptions have been attended (APPENDIX **2**). In this new model, (Table **4**) H_{int}, pVV and pHV are the only variables required to estimate forest biomass:

$\log_{10}AGB$ (Ton/ha) = -6.54 + 6.94 ($\log_{10} H_{int}$) + 0.03 pHV – 0.03 pVV (4)

Standard deviation of model parameters ($\pm$ 2.26), ($\pm$ 0.54) , ($\pm$ 0.02), ($\pm$ 0.01) respectively

This model can be algebraically re-arranged as:

$\log_{10}AGB$ (Ton/ha) = -6.54 + 6.94 ($\log_{10}H_{int}$) + 0.03 (pHV – pVV)   (5)
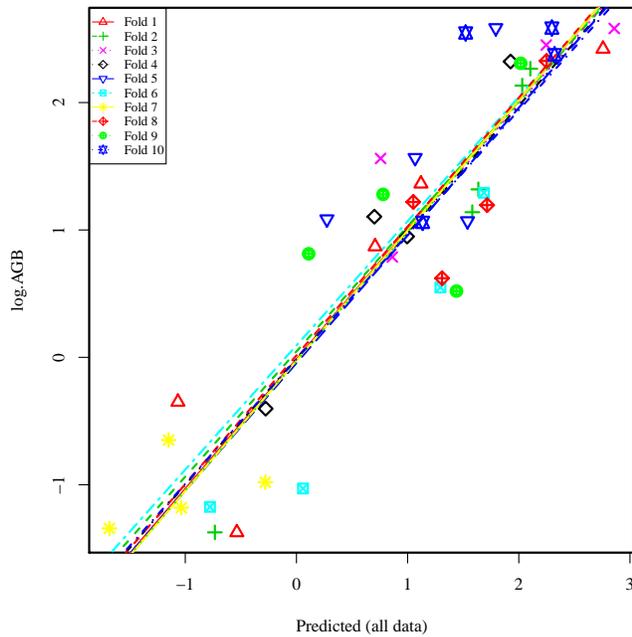
Fig. (11). Cross validation k-fold (k=10) fits of the linear model $\log_{10}AGB = \log \alpha + b \log_{10}H_{int} + \beta5 \, (pHV - pVV)$.

The reanalysis of this model using $(pHV - pVV)$ as a compounded variable produces the following model (Table **4**, Fig. **10**):

$\log_{10}AGB$ (Ton/ha) = -6.68 + 6.94 $(\log_{10} H_{int})$ + 0.03 $(pHV - pVV)$ (6)

with standard deviation of model parameters ($\pm$ 0.60), ($\pm$ 0.53), and ($\pm$ 0.01) respectively.

This new model is still highly significant (p < 0.0001) with an adjusted R^2 = 0.82, and RMSE=0.544. The normalized RMSE with respect to the $\log_{10}AGB$ maximum (2.5855) and minimum (-1.3743) data points is 13.7%.

There is no need to force the intercept through zero because the intercept value is already very close to zero. Residuals were normal (Shapiro Wilks test p=0.70). Diagnostic plots to verify whether model assumptions have been attended can be seen in APPENDIX 3. The difference between pHV and pVV showed a minor but significant influence on forest biomass (R^2 of models with and without $(pHV - pVV)$ are respectively 0.82 and 0.78). This indicates that H_{int} explains most of the variation in forest biomass.

The k-fold (k=10) cross-validation algorithm applied to the final model estimated an overall mean square of 0.335, with a root overall mean square=0.579. The similarity between this estimate and the model's RMSE estimate suggests that the cross-validation analysis estimates models very similar to the model containing the entire data set. Fig. (**11**) shows a plot of the 10-fold fit obtained from the cross-validation analysis. The AIC values for models using pHV and pVV independently and $(pHV - pVV)$ are respectively 74.95 and 72.95. Comparing both models with the function anova of R did not result in a significant difference. Although the gain in fit when using the compounded variable $(pHV - pVV)$ is small and non-significant, reducing the number of parameters in a model represents, in our view,

an improvement for interpretation. Thus, this model with fewer parameters was chosen for discussion. The statistical model indicates that pHH has no statistical influence on the $\log_{10}AGB$. This result contrasts somewhat with Fig. (**6**), which shows that values of the parameter pHH are lower and less variable in the class FE than the other classes, indicating that variation in pHH is associated with variation in some forest features, such as forest exploration in this case.

Map of the Forest Above-Ground Biomass

H_{int} and $(pHV - pVV)$ are the variables used to estimate the per-pixel biomass according to the linear model fitted to data. For the chosen area, the distribution of each variable was obtained with ENVI. These distributions showed some pixels that considerably exceeded expectations. For example, few pixel values for H_{int} were greater than 40 meters, which represents forest heights much greater than the average Amazon Forest [69]. Several factors (*e.g.*, data gathering or processing errors, the presence of rare and very large trees, and other unknown factors) may have caused such unexpected large values. If some of these larger values are caused by errors, forest biomass can be potentially overestimated. Since, at this point, we can only speculate which are such causal factors, we have decided to use the following algorithm to avoid forest biomass overestimation: 1) Using the per pixel biomass distribution, we have identified the values above which lies 5% of the distribution. 5% is conventionally used as an acceptable cutoff value for type I error, *i.e.*, the probability of rejecting a true hypothesis. By using this criterion, the upper 5% of the distribution corresponds to: H_{int} >20.5m and $(pHV - pVV) >$ 49. 2) Pixels greater than these cutoff values were taken to be equal to the corresponding cutoff values. As such, the greatest H_{int} was 20.5 m and $(pHV - pVV)$ was 49. Recent studies [69] estimated average forest heights in most of the Amazon region, including the Paragominas region, shorter than the upper limit of 20.5m set here. The vegetation biomass map is shown in Fig. (**12**). The average vegetation biomass was 73 Ton/ha, ranging from pixels with zero biomass in non-forested areas, and pixels with biomass greater than 600 Ton/ha observed in regions of dense forest. The forest biomass of the selected area within the red rectangle (853 ha) was estimated in 115,595 Tons, with an average of 135 Ton/ha (Fig. **12**).

DISCUSSION

Our results suggest two features that make this dual-band InSAR methodology adequate to estimate forest biomass and detect forest degradation: 1) the ability to detect forest exploration and degradation by combining estimates of the interferometric height H_{int} with the backscatter polarization pHH, and 2) the possibility of measuring forest biomass through a reliable relationship with the interferometric height H_{int} and the compounded variable $(pHV - pVV)$.

Backscatter - Forest Exploration and Degradation

Several attempts have been made to take advantage of airborne radar information in forest monitoring [21, 31, 32, 40, 41, 45, 70-72]. Forest structure – *i.e.*, tree number, size and spatial distribution – has the potential to change how P-band microwaves are scattered [42, 73]. In addition, different patterns of X-band reflection could be explained by how

Map of Biomass Paragominas
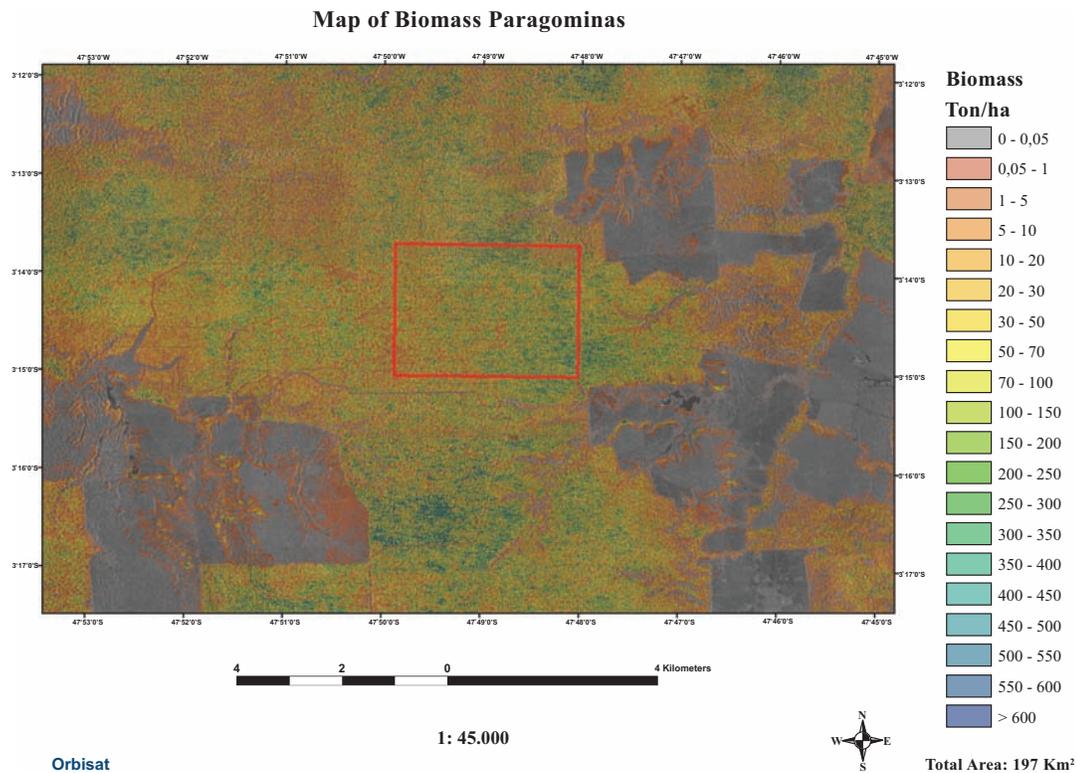
Fig. (12). Map of the forest biomass of an area of 197 km2 in Paragominas. Darker areas correspond to a greater biomass. The estimated biomass for the sampled area (limited by the rectangle) is 115,595 Tons, with an average of 135 Ton/ha. A closer analysis of this map can be used to understand details of the deforestation process, such as the areas where forest disturbance predominately occurs.

forest canopies modify X-band backscatter. Thus, it is theoretically possible to classify an entire landscape based solely on how radar attributes interact with different land use/land cover classes.

The classification algorithm seems to have been trained to classify the vegetation from FD to SS based on, among other factors, increasing values of the variable *xHH*, which measures the X-band backscatter (Fig. **6**). From this, we can hypothesize that denser forests present either: 1) smoother canopies than early succession vegetation (SS), reflecting a greater proportion of X-bands away from the source airplane; 2) more absorbing canopies than early succession vegetation (SS), retaining an increasing proportion of arriving X-bands; or 3) canopies that are much more heterogeneous than early succession vegetation (SS), causing multiple X-band scattering at the canopy level and preventing reflected X-bands from returning to the source airplane.

Forest exploration can be quite destructive due to the impact of falling trees in the vegetation as well as the effect of creating dragging tracks to remove these trees from the area [74, 75]. Even when sustainable management is employed, forest exploration can cause substantial damage [74, 75]. Thus, dense forest exploration likely reduces the size and density of trees from which P-band microwaves reflect, thereby reducing backscatter levels. Based on the information provided by the field crew, the classification algorithm interpreted lower *pHH* backscatter levels as an indication that a given area has been explored. These features likely result from a lower complexity in the forest structure found in FE plots caused by the above-described

exploration process. Our results suggest that the radar variable *pHH*, despite presenting limitations to estimate tropical forest biomass (see below), may in fact be an excellent indicator of forest exploration and degradation [26, 44, 45] that end up reducing large tree density in an area, resulting in a less effective double bounce process [42].

Above Ground Biomass (*AGB*) Estimation

Our results confirm that, among all InSAR radar variables, H_{int} was the most relevant radar-derived variable explaining forest biomass, usually accounting for more than 80% of the variance in forest biomass [31, 40, 41], likely a consequence of the close relationship between H_{int} and tree height, which is a variable commonly used in allometric models to predict tree biomass [8-10, 74, 76]. Thus, a close relationship between forest interferometric height and forest biomass is not surprising.

The slight, but significant, improvement in predicting forest biomass with the inclusion of backscatter variables in the statistical model merits some considerations. Although, radar backscatter correlates with forest biomass until it saturates [26, 31, 44, 45], forest biomass affects radar polarizations differently. Some studies report that in the P-band case, *pHV* and *pHH* better correlates with forest biomass components – *e.g.*, *DBH*, tree height, or basal area – than *pVV* [26, 33, 34, 77]. In a study using several L-band polarizations, the *LHV* polarization was shown to better estimate forest biomass in the Colombian Amazon than alternative L-band polarizations [58]. In addition, because different bands [29] and different polarizations of a band [34, 73] interact differently with different forest strata – *e.g.*, P-

bands penetrate the forest completely, X-band cannot penetrate the forest, and C- and L-bands penetrate the forest in different degrees, – compounded variables such as the ratios *pHV/CHV* and *LHV/CHV* [29], *LHH/LVV* and *pHH/pVV* and *pHV/pHH* ratios [78], the L phase difference *HH–VV* [79], and the combination of *LHH* and the ratio *pHH/pVV* [26] showed some slight improvement in forest biomass estimation over single radar variables. The model obtained here indicates that the forest biomass also depends on the compounded variable (*pHV – pVV*). Everything else equal, increasing *pHV* results in a greater *AGB*. And, everything else equal, decreasing *pVV* results in a greater *AGB*. Larger forests seemingly increase the rate of cross polarization (producing a greater *pHV*) possibly due to micro interactions with larger tree trunks and branches, while decrease (with respect to *pHV*) the vertical P-band backscatter.

Our model contrasts somewhat with previous radar-based models where *pHH* showed a minor but significant importance in predicting forest biomass [40, 41]; here *pHH* does not contribute to predicting forest biomass. The reason for this behavior likely has multiple causes. Studies that obtained such significant influence of *pHH* on forest biomass either included cleared land plots, such as pastures [41], or eucalyptus plantations [40], which cause a substantial reduction in the P-band backscatter when compared to dense tropical forests. We neither used cleared land plots nor commercial forests and, thus, the differences between forest classes with respect to *pHH* were not large. The only exception was FE. FE and FS showed similar H_{int}, but FE presented a significant lower *pHH* than all other forest classes. In addition, H_{int} of SS was much smaller than that of FE, but the value of *pHH* in FE was still smaller than that of SS. This causes a disconnection between forest biomass, H_{int}, and *pHH*. Only if *pHH* and H_{int} were well correlated could one expect a larger contribution from *pHH* to forest biomass than observed here.

With the clues obtained in this analysis in light of the results already available in the literature, we have reached a point where manipulative experiments are becoming increasingly necessary to quantitatively disentangle the effects of forest structure on backscatter parameters. For example, the canopy of some forest plots could be partially removed in different degrees, and then P-band polarizations could be compared among experimental plots. Nonetheless, our results confirm that a great deal of information about forest biomass and structure can be obtained by combining different radar attributes.

The statistical model derived from this experiment contrasts quite a lot with the model described in [41]. This contrast mainly results from the model in [41] being a linear model with biomass in a non-transformed scale, while biomass in our model follows the general allometric equation $AGB = a\,H_{int}^{b}$. Part of this between-model contrast may also be the result of a smaller sample size (n=19) in [41] than used here (n=42), making it harder to detect non-linearities in the data. However, differences in forest physiognomies – the study in [41] was conducted in the Tapajós region - may have added to the differences in model fitting. By comparing both models, strong differences in biomass estimates of early

successional forests can be observed: [41] estimates a much greater biomass in younger/smaller forests than estimated by our model, even if (*pHV – pVV*) < 0 (in our model) and considering *pHH*=0 (in [41]'s model) (Table **5**). The DTM and DSM vertical imprecision of the present data set can be up to $\pm$ 1.5m (1 σ). Thus, estimating the vegetation biomass becomes relatively more imprecise in shorter vegetation. This may constitute an additional contribution to reduce the biomass estimates of early successional/short vegetation in our model. However, because forest biomass estimation is also modulated by both *pHH* in [41] and (*pHV – pVV*) here, model comparison (Table **5**) must be interpreted with caution.

The total forest biomass of an area is the sum of each individual plant size/biomass found in this area. However, tree sizes become increasingly heterogeneous as forests grow; fewer larger trees contribute disproportionately to the total biomass, just like they contribute disproportionately to H_{int} estimation. This is consistent with current expectations of plant populations and communities predicting the development of plant size hierarchies as plant competition for light increases [80], which is the case in mature forests when compared to earlier succession forests. The allometric model underlying our analysis is based on the equation $AGB = a\,H_{int}^{b}$, which was borrowed from a single-tree biomass equation. As a result, the model implicitly assumes that, as forests grow, a homogeneous tree stand forms the forest canopy and determines H_{int}, which is not the case, as our data clearly demonstrates (Fig. **8**). This can potentially overestimate the biomass of forests taller than the forests assessed in this study, where size hierarchies might be even more pronounced – *i.e.*, the forest canopy being formed by fewer larger trees. The analysis of data that include forests taller than those found in Paragominas is necessary to obtain more general statistical models suitable to confidently estimate the biomass in a broader range of forest physiognomies.

A potential caveat of our *AGB* analysis relates to the allometric corrections applied to account for the forest growth between flight and forest inventories. Unfortunately, temporal gaps in datasets, such as the one in this study are common (*e.g.*, [41]). Secondary forests and secondary succession vegetation grow quite rapidly in the tropics. Forests grow non-linearly and growth tends to behave asymptotically after 40-50 years of age, and so, relative growth rates of early succession forests are much higher than those observed in mature forests. [61]. Therefore, not accounting for growth along the three years elapsed between flight and forest inventories would result in an overestimation of plot biomass in FE, FS and SS with respect to FD, producing systematic errors in the statistical model. The main question is whether the growth model used for such adjustment/correction in size and biomass is adequate. To date, the model described in [61] is the most comprehensive growth model for secondary forest growth in the Amazon region, for *it addresses all forest parameters, commonly incorporated in quantitative descriptions of forest growth* [61]. Having in mind that flight and field data taken within a short timeframe would be ideal for analysis, applying such corrections is a valid strategy to reduce potential biases in the available dataset.

Conservative approaches are recommended to estimate

Table 5.   **Comparison between the forest biomass estimated with the equation developed here, and the linear equation obtained in [41] as a function of H_{int} in meters. Simulations include estimates with ($pHV – pVV$)=0, ($pHV – pVV$)=10, and ($pHV – pVV$)=-10. Note that, for this comparison, the backscatter parameter in [41] equation was taken to be zero. Since this is a positive parameter, and that it adds to the H_{int} term in the biomass estimate, the biomass estimate with zero backscatter is an underestimate. Nonetheless, the table clearly shows the differences in model behavior.**

H_{int} (m)	Biomass (Ton/ha) *Model with* ($pHV – pVV$)=0	Biomass (Ton/ha) *Model with* ($pHV – pVV$)=10	Biomass (Ton/ha) *Model with* ($pHV – pVV$)=-10	Biomass (Ton/ha) [41] *model with zero backscatter*
1	~0	~0	~0	58.83
5	0.02	0.04	0.01	114.27
10	0.39	4.9	1.23	183.57
15	40.77	81.34	20.43	252.87
20	299.32	597.23	150.02	322.17
21	419.74	837.49	210.36	336.03
22	579.41	1156	290.39	349.89

the forest biomass for REDD projects (http://www.v-c-s.org/methodology_rmm.html). The chosen algorithm for limiting the upper value of variables to the 5% cutoff constitutes a conservative approach in that it avoids biomass overestimation when outliers result from data problems instead of true large trees, something we cannot differentiate at this point. This is particularly relevant when working in a log scale, where one unit in a log scale roughly corresponds to one order of magnitude in biomass.

CONCLUSION

Several remote sensing technologies are being developed to provide accurate estimates of forest biomass to guarantee that REDD can effectively reduce carbon emissions from deforestation and forest degradation. X/P-band InSAR combines the potential for wall-to-wall data acquisition in very large areas and estimation of forest biomass in conditions - such as with the presence of cloud cover, and dense vegetation – that other remote sensing technologies face difficulties to deliver. We confirmed the results of previous studies [31, 40, 41] demonstrating that, when H_{int} is available, forest biomass estimation is improved substantially. This can overcome the known limitation of using P-band backscatter alone for biomass estimation, *i.e.*, signal saturation above a certain biomass threshold. We also showed that the P-band backscatter alone can be used to obtain important information about changes in forest structure, such as those resulting from forest exploration. This can become quite desirable for REDD projects themselves as well as for countries and states or provinces establishing country-wide monitoring systems as part of their *readiness* [81] activities to receive REDD projects. To date, only few studies using X- and P-band InSAR have been conducted in the Amazonian rainforest. It has been shown that forests across the globe can possess different allometric relationships [56, 82], which might prevent the application of a single functional relationship between radar attributes and forest biomass across the entire Amazon region. We, thus, propose that the development of a more general protocol by which forest biomass can be estimated for contrasting forest physiognomies and ecosystems of the Amazon and other

tropical regions of the world using airborne InSAR data is a research priority.

CONFLICT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

APPENDIX 1

Confusion Matrix – Ground Truth x Radar Classified

```
Overall Accuracy = (36561/39855)  91.7350%
Kappa Coefficient = 0.9026

                      Ground Truth (Pixels)
        Class      Exposed soil  Abandoned Past.    FE    Clean Past.    SS
Unclassified            0              0             0         0          0
Exposed soil          2902             0             0         0          0
Abandoned Past.       1109           3769            0        39          0
FE                      0              0           2676        0          0
Clean Past.            73             25             0       3165         0
SS                     200             0             0         0        8470
FD                      0              0             0         0          0
FS                      0              0             0         0          0
Water-Wet               0              0             0         0          0
    Total            4284           3794          2676      3204       8470

                      Ground Truth (Pixels)
        Class          FD            FS        Water-Wet     Total
Unclassified            0              0             0          0
Exposed soil            0              0             0        2902
Abandoned Past.         0              0             0        4917
FE                     885            758            0        4319
Clean Past.             0              0             0        3263
SS                      0             12             0        8682
FD                    8435             4             0        8439
FS                     189           4833            0        5022
Water-Wet               0              0           2311       2311
    Total             9509           5607          2311      39855

                      Ground Truth (Percent)
        Class      Exposed soil  Abandoned Past.    FE    Clean Past.    SS

Unclassified          0.00           0.00          0.00      0.00       0.00
Exposed soil         67.74          99.34          0.00      0.00       0.00
Abandoned Past.      25.89           0.00          0.00      1.22       0.00
FE                    0.00           0.00         100.00     0.00       0.00
Clean Past.           1.70           0.66          0.00     98.78       0.00
SS                    4.67           0.00          0.00      0.00      100.00
FD                    0.00           0.00          0.00      0.00       0.00
FS                    0.00           0.00          0.00      0.00       0.00
Water-Wet             0.00           0.00          0.00      0.00       0.00
    Total           100.00         100.00        100.00    100.00     100.00
```

Appendix 1. Contd….

	Ground Truth (Percent)			
Class	FD	FS	Water-Wet	Total
Unclassified	0.00	0.00	0.00	0.00
Exposed soil	0.00	0.00	0.00	7.28
Abandoned Past.	0.00	0.00	0.00	12.34
FE	9.31	13.52	0.00	10.84
Clean Past.	0.00	0.00	0.00	8.19
SS	0.00	0.21	0.00	21.78
FD	88.71	0.07	0.00	21.17
FS	1.99	86.20	0.00	12.60
Water-Wet	0.00	0.00	100.00	5.80
Total	100.00	100.00	100.00	100.00

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Exposed soil	0.00	32.26	0/2902	1382/4284
Abandoned Past.	23.35	0.66	1148/4917	25/3794
FE	38.04	0.00	1643/4319	0/2676
Clean Past.	3.00	1.22	98/3263	39/3204
SS	2.44	0.00	212/8682	0/8470
FD	0.05	11.29	4/8439	1074/9509
FS	3.76	13.80	189/5022	774/5607
Water-Wet	0.00	0.00	0/2311	0/2311

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Exposed soil	67.74	100.00	2902/4284	2902/2902
Abandoned Past.	99.34	76.65	3769/3794	3769/4917
FE	100.00	61.96	2676/2676	2676/4319
Clean Past.	98.78	97.00	3165/3204	3165/3263
SS	100.00	97.56	8470/8470	8470/8682
FD	88.71	99.95	8435/9509	8435/8439
FS	86.20	96.24	4833/5607	4833/5022
Water-Wet	100.00	100.00	2311/2311	2311/2311

Most classification inacuracies are concentrated in mis-classifying exposed soil as pastures. Such inacuracies are not relevant for this work, as we did not use these classes in our analyses.
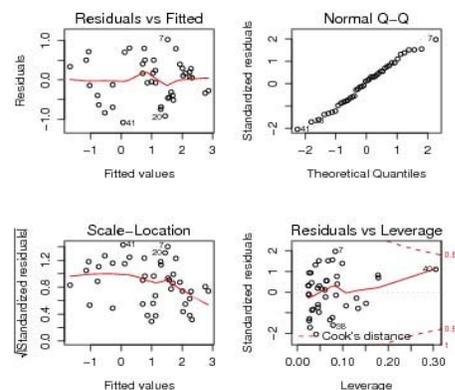
APPENDIX 2

Diagnostic plots of Model 1 –

$$\log_{10}AGB = \log(\alpha) + \beta_0 \log_{10}H_{int} + \beta_2\, pVV + \beta_3\, pHV$$
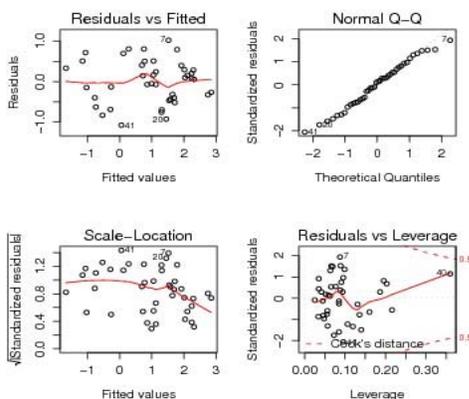
APPENDIX 3

Diagnostic plots of Model 2

$$LOG_{10}AGB = LOG(A) + B_0\, LOG_{10}H_{INT} + B_5(PHV - PVV)$$

REFERENCES

[1] Saatchi SS, Harris NL, Brown S, *et al*. A Benchmark map of forest carbon stocks in tropical regions across three continents. Proc Natl Acad Sci USA 2011; 108(24): 9899-904.

[2] Friedlingstein P, Houghton RA, Marland G, *et al.* Update on CO2 emissions. Nat Geosci 2010; 3: 811-2.

[3] Fearnside PM. Greenhouse gas emissions from land-use change in Brazil's Amazon region. In: Lal R, Kimble JM, Stewart B, Eds. Global climate change and tropical ecosystems advances in soil science. Boca Raton, Florida, USA: CRC Press 2000, pp. 231-49.

[4] Meyfroidt P, Rudel TK, Lambin EF. Forest transitions, trade, and the global displacement of land use. Proc Natl Acad Sci USA 2010; 107(49): 20917-22.

[5] Asner GP. Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. Environ Res Lett 2009; 4: 1-11.

[6] Gibbs HK, Brown S, Niles JO, Foley JA. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environ Res Lett 2007; 2: 045023.

[7] Olander LP, Gibbs HK, Steininger M, Swenson JJ, Murray BC. Reference scenarios for deforestation and forest degradation in support of REDD: a review of data and methods. Environ Res Lett 2008; 3(2): 025011.

[8] Brown S. Measuring carbon in forests: current status and future challenges. Environ Pollut 2002; 116: 363-72.

[9] Brown S, Gaston G. Use of forest inventories and geographic information systems to estimate biomass density of tropical forests: applications to tropical Africa. Environ Monit Assess 1995; 38: 157-68.

[10] Brown S, Gillespie AJR, Lugo AE. Biomass estimation methods for tropical forest with application to forest inventory data. Forest Sci 1989; 35: 881-902.

[11] Chave J, Andalo C, Brown S, *et al.* Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 2005; 145(1): 87-99.

[12] DeFries R, Achard F, Brown S, *et al*. Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. Environ Sci Policy 2007; 10: 385-94.

[13] INPE. Monitoramento da floresta Amazonica Brasileira por satellite. São José dos Campos; 2010. Available from: http://www.obt.inpe.br/prodes/ [updated 2010; cited 2010, 10 December].

[14] Asner GP. Cloud cover in Landsat observations of the Brazilian Amazon. Int J Remote Sens 2001; 22: 3855-62.

[15] Foody GM, Boyd DS, Cutler MEJ. Predictive relations of tropical forest biomass from Landset TM data and their transferability between regions. Remote Sens Environ 2003; 85: 463-74.

[16] Thenkabail PS, Enclona EA, Ashton MS. Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rain forests. Remote Sens Environ 2004; 90: 23-43.

[17] Drake JB, Knox RG, Dubayah RO, *et al.* Above-ground biomass estimation in closed-canopy neotropical forests using lidar remote sensing: factors affecting the generality of relationships. Glob Ecol Biogeogr 2003; 12: 147-59.

[18] Lefsky MA, Cohen WB, Parker GG, Harding DJ. Lidar remote sensing for ecosystem studies. Bioscience 2002; 52(1): 19-30.

[19] Lefsky MA, Hudak AT, Cohen WB, Acker SA. Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest. Remote Sens Environ 2005; 95: 532-48.

[20] Sexton JO, Bax T, Siqueira P, Swenson JJ, Hensley S. A comparison of LiDAR, radar, and field measurements for estimating canopy height in pine and hardwood forests of southeastern North America. For Ecol Manage 2009; 257: 1136-47.

[21] Treuhaft RN, Gonçalves FG, Drake JB, *et al*. Biomass estimation in a tropical wet forest using Fourier transforms of profiles from lidar or interferometric SAR. Geophys Res Lett 2010; 37: L23403.

[22] Lucas R, Armston J, Fairfax R, *et al.* An Evaluation of the ALOS PALSAR L-Band Backscatter — Above Ground Biomass Relationship Queensland, Australia: Impacts of Surface Moisture Condition and Vegetation Structure. IEEE J Sel Top Appl Earth Observ 2010; 3: 576-93.

[23] Santoro M, Beer C, Cartus O, *et al*. Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. Remote Sens Environ 2011; 115(2): 490-507.

[24] Lee J-S, Ainsworth TL, Wang Y, Chen K-S, Chih TW. Monitoring tree farms and coastal environments using RADARSAT-2 PolSAR

data. Geoscience and Remote Sensing Symposium (IGARSS), IEEE International 2010; 25-30 Jul 2010; Honolulu, Hawaii 2010.

[25] Holopainen M, Haapanen R, Karjalainen M, *et al*. Comparing accuracy of airborne laser scanning and TerraSAR-X radar images in the estimation of plot-level forest variables. Remote Sens 2010; 2: 432-45.

[26] Dobson MC, Ulaby FT, Le Toan T, Beaudoin A, Kasischke ES, Christensen N. Dependence of radar backscatter on conifer forest biomass. IEEE Trans Geosci Remote Sens 1992; 30: 412-5.

[27] Le Toan T, Quegan S, Woodward I, Lomas M, Delbart N, Picard C. Relating radar remote sensing of biomass to modeling of forest carbon budgets. Clim Change 2004; 76: 379-402

[28] Patenaude G, Hill RA, Milne R, *et al*. Quantifying forest above ground carbon content using LiDAR remote sensing. Remote Sens Environ 2004; 93: 368-80.

[29] Ranson KJ, Sun G. Mapping biomass of a northern forest using multifrequency SAR data. IEEE Trans Geosci Remote Sens 1994; 32: 388-96.

[30] Rosenqvist A, Shimada M, Igarashi T, *et al*. Support to multi-national environmental conventions and terrestrial carbon cycle science by ALOS and ADEOS-II–the Kyoto and carbon initiative. Geoscience and Remote Sensing Symposium, 2003, Proc. IEEE International 2003.

[31] Saatchi SS, Marlier M, Chazdon RL, Clark DB, Russel AE. Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. Remote Sens Environ 2010. Doi.10.1016/j.res.2010.07.015.

[32] Shimada M, Rosenqvist A, Watanabe M, *et al*. The polarimetric and interferometric potential of ALOS PALSAR. POL in SAR Frascati, Italy 2005.

[33] Wang Y, Davis FW, Melack JM, Kasischke ES, Christensen Jr NL. The effects of changes in forest biomass on radar backscatter from tree canopies. Int J Remote Sens 1995; 16(3): 503-13.

[34] Wu S-T, Sader SA. Multipolarization SAR data for surface feature delineation and forest vegetation characterization. IEEE Trans Geosci Remote Sens 1987; GE-25(1): 67-76

[35] Castilho CV, Magnusson WE, Araujo RNO, *et al.* Variation in aboveground tree live biomass in a central Amazonian Forest: Effects of soil and topography. For Ecol Manage 2006; 234: 85-96.

[36] Fearnside P. Greenhouse gases from deforestation in Brazilian Amazonia: Net committed emissions. Clim Change 1997; 35: 321-60.

[37] Houghton R, Skole D, Nobre C, Hackler J, Lawrence K, Chomentowski W. Annual fluxes of carbon from deforestation and regrowth in the Brazilian Amazon. Nature 2000; 403(20): 301-4.

[38] Neeff T, Graça PMA, Dutra LV, Freitas CC. Carbon budget estimation in Central Amazonia: successional forest modeling from remote sensing data. Remote Sens Environ 2005; 94: 508-22.

[39] Saatchi SS, Houghton RA, Dos Alvalá SRC, Soares JV, Yu Y. Distribution of aboveground live biomass in the Amazon basin. Glob Change Biol 2007; 13: 816-37.

[40] Gama FF, Santos JR, Mura JC. Eucalyptus biomass and volume estimation using interferometric and polarimetric SAR data. Remote Sens 2010; 2: 939-56.

[41] Neeff T, Dutra LV, Santos JR, Freitas CC, Araujo LS. Tropical forest measurement by interferometric height modeling and P-band radar backscatter. Forest Sci 2005; 51(6): 585-94.

[42] Lucas RM, Cronin N, Moghaddam M, *et al*. Integration of radar an Landsat-derived foliage projected cover for woody regrowth mapping, Queensland, Australia. Remote Sens Environ 2006; 100: 388-406.

[43] Caicoya AT, Kugler F, Papathanassiou K. Biomass estimation as a function of vertical forest structure and forest height. Potential and limitations for Radar Remote Sensing. EUSAR 2010; Aachen, Germany 2010.

[44] Santos JR, Araujo LS, Kuplich TM, *et al.* Tropical biomass and its relationship with P-band SAR data. Revista Brasileira de Cartografia 2006; 58(1): 37-42.

[45] Santos JR, Freitas CC, Araujo LS, *et al*. Airborne P-band SAR applied to the aboveground biomass studies in the Brazilian tropical rainforest. Remote Sens Environ 2003; 87: 482-93.

[46] Centro de Pesquisa Agropecuária do Trópico Úmido: Laboratório de Climatologia. Normais climatológicas de Paragominas no período de 1980 a 1988. Belém: Embrapa 1986 Contract No.: Document Number.

[47] Richards J. Remote sensing digital image analysis: an introduction. Berlin: Springer Verlag 1999.

[48] Jensen JR. Introductory digital image processing. London, UK: Prentice-Hall Int 1996.

[49] Swain PH, Davis SM. Remote sensing: the quantitative approach New York: McGraw-Hill 1978.

[50] Müllerová J. Use of digital aerial photography for sub-alpine vegetation mapping: A case study from the Krkonoše Mts, Czech Republic. Plant Ecol 2004; 175: 259-72.

[51] Thomas V, Treitz P, Jelinski D, Miller J, Lafleur P, McCaughey JH. Image classification of a northern peatland complex using spectral and plant community data. Remote Sens Environ 2002; 84: 83-99.

[52] Dutra LV, Correira AH, Mura JC, Santos JR, Elmiro MAT, Freitas CC, Eds. Avaliação das imagens polarimétricas da simulação MAPSAR para classificação de uso/ocupação do solo na região da Floresta Nacional do Tapajós. Simpósio Brasileiro de Sensoriamento Remoto. Brazil: Florianópolis 2007.

[53] Lee J-S, Grunes MR, Pottier E. Quantitative comparison of classification capability: fully polarimetric versus dual and single polarization SAR. IEEE Trans Geosci Remote Sens 2001; 39(11): 2343-51.

[54] Molina I, Fernández S, Ormeño S, Papi F, González E. Evaluation of ALOS-PALSAR polarimetric modes capabilities for landcover monitoring and mapping. In: SP-664 E, Ed. ALOS PI 2008 Symposium; 2009; Island of Rhodes, Greece 2009; pp. 1-6.

[55] Neto PS, Brena DA. Inventário Florestal. Cutitiba PR, Brazil; 1997.

[56] Feldpausch TR, Banin L, Phillips OL, *et al*. Height-diameter allometry of tropical forest trees. Biogeosciences 2011; 8(5): 1081-106.

[57] Higgins MA, Ruokolainen K. Rapid tropical forest inventory: a comparison of techniques based on inventory data from western Amazonia. Conserv Biol 2004; 18: 799-811.

[58] Hoekman DH, Quiñones MJ. Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Colombian Amazon. IEEE Trans Geosci Remote Sens 2000; 38(2): 685-96.

[59] Vieira ICG, Almeida AS, Davidson EA, *et al*. Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazonia. Remote Sens Environ 2003; 87: 470-81.

[60] Gonçalves FG, Santos JR. Composição florística e estrutura de uma unidade de manejo florestal sustentável na Floresta Nacional do Tapajós, Pará. Acta Amaz 2008; 38(2): 229-44.

[61] Neeff T, Santos JR. A growth model for secondary forest in Central Amazonia. For Ecol Manage 2005; 216(1-3): 270-82.

[62] Chave J, Coomes D, Jansen S, *et al*. Towards a worldwide wood economics spectrum. Ecol Lett 2009; 12: 351-66.

[63] Zanne AE, Lopez-Gonzalez G, Coomes DA, *et al.* Global wood density database. Dryad. Identifier. 2009. Available from: http://hdl.handle.net/10255/dryad.235 [updated 2009; cited 2011].

[64] Huxley JS. Problems of relative growth. New York: MacVeagh 1932.

[65] Mette T, Papathanassiou K, Hajnsek I, Pretzsch H, Biber P, Eds. Applying a common allometric equation to convert forest height from Pol-InSAR data to forest biomass. Geoscience and Remote Sensing Symposium, IGARSS '04; 2004.

[66] Zianis D, Mencuccini M. Aboveground biomass relationships for beech (*Fagus moesiaca* Cz.) trees in Vermio Mountain, Northern Greece, and generalised equations for *Fagus sp*. Ann For Sci 2003; 60: 439-48.

[67] Venables WN, Ripley BD. Modern applied statistics with S. New York: Springer-Verlag Inc 2002.

[68] R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing 2008. Available from: http://www.R-project.org/. [updated 2008].

[69] Lefsky MA. A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. Geophys Res Lett 2010; 37(L15401): 1-5.

[70] Liu D, Du Y, Sun G, Yan WZ, B-I W. Analysis of insar sensitivity to forest structure based on radar scattering model. PIER 2008; 84: 149-71.

[71] Solberg S, Astrup R, Bollandsås OM, Næsset E, Weydahl DJ. Deriving forest monitoring variables from X-band InSAR SRTM height. Can J Rem Sens 2010; 36: 68-79.

[72]   Treuhaft RN, Siqueira PR. Vertical structure of vegetated land surfaces from interferometric and polarimetric radar. Radio Sci 2000; 35(1): 141-77.

[73]   Kasischke ES, Christensen Jr NL, Bourgeau-Chavez LL. Correlating radar backscatter with components of biomass in Loblolly pine forest. IEEE Trans Geosci Remote Sens 1995; 33: 643-59.

[74]   Uhl C, Kauffman JB. Deforestation effects on fire susceptibility and the potencial response of tree species in the rain forest of the eastern Amazon. Ecology 1990; 90: 437-49.

[75]   Veríssimo A, Barreto P, Tarifa R, Uhl C. Extraction of a high-value natural source from Amazon: the case of mahogany. For Ecol Manage 1995; 55: 169-99.

[76]   Uhl C, Buschbacher R, Serrão EAS. Abandoned pastures in eastern Amazonia - I: Patterns of plant succession. J Ecol 1988; 76: 663-81.

[77]   Le Toan T, Beaudoin A, Riom J, Guyon D. Relating forest biomass to SAR data. IEEE Trans Geosci Remote Sens 1992; 30: 403-11.

[78]   Freeman A, Durden SL. A three-component scattering model for polarimetric SAR data. IEEE Trans Geosci Remote Sens 1994; 36(3): 963-73.

[79]   Gonçalves FG, Santos J, Treuhaft RN. Stem volume of tropical forests from polarimetric radar. Int J Remote Sens 2011; 32(2): 503-22.

[80]   Weiner J. Asymmetric competition in plant populations. Trends Ecol Evol 1990; 5: 360-4.

[81]   Barr C, Dermawan A, Purnomo H, Komarudin H. Readiness for REDD: Financial governance and lessons from Indonesia's Reforestation Fund (RF). CIFOR Infobriefs. Bogor: Center for International Forestry Research (CIFOR); 2009 November. Report No. 20.

[82]   Nogueira EM, Nelson BW, Fearnside PM, França MB, Oliveira ACA. Tree height in Brazil's 'arc of deforestation: Shorter trees in south and southwest Amazonia imply lower biomass. For Ecol Manage 2008; 255: 2963-72.

.