Global Change Biology (2007) **13,** 816–837, doi: 10.1111/j.1365-2486.2007.01323.x

Distribution of aboveground live biomass in the Amazon basin

S. S. SAATCHI*, R. A. HOUGHTON†, R. C. DOS SANTOS ALVALÁ‡, J. V. SOARES‡ and Y. YU*

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, †Woods Hole Research Center, Woods Hole, MA, USA, ‡Instituto Nacional de Pesquisas Espaciais – INPE, São José dos Campos, SP, Brazil

Abstract

The amount and spatial distribution of forest biomass in the Amazon basin is a major source of uncertainty in estimating the flux of carbon released from land-cover and land-use change. Direct measurements of aboveground live biomass (AGLB) are limited to small areas of forest inventory plots and site-specific allometric equations that cannot be readily generalized for the entire basin. Furthermore, there is no spaceborne remote sensing instrument that can measure tropical forest biomass directly. To determine the spatial distribution of forest biomass of the Amazon basin, we report a method based on remote sensing metrics representing various forest structural parameters and environmental variables, and more than 500 plot measurements of forest biomass distributed over the basin. A decision tree approach was used to develop the spatial distribution of AGLB for seven distinct biomass classes of lowland old-growth forests with more than 80% accuracy. AGLB for other vegetation types, such as the woody and herbaceous savanna and secondary forests, was directly estimated with a regression based on satellite data. Results show that AGLB is highest in Central Amazonia and in regions to the east and north, including the Guyanas. Biomass is generally above 300 Mg ha^{-1} here except in areas of intense logging or open floodplains. In Western Amazonia, from the lowlands of Peru, Ecuador, and Colombia to the Andean mountains, biomass ranges from 150 to 300 Mg ha^{-1}. Most transitional and seasonal forests at the southern and northwestern edges of the basin have biomass ranging from 100 to 200 Mg ha^{-1}. The AGLB distribution has a significant correlation with the length of the dry season. We estimate that the total carbon in forest biomass of the Amazon basin, including the dead and belowground biomass, is 86 Pg C with ± 20% uncertainty.

Keywords: Amazon, biomass, climate, rainforest, remote sensing

Received 15 July 2005; revised version received 16 February 2006 and accepted 13 July 2006

Introduction

The distribution of forest biomass over tropical forests is uncertain. Current estimates for the forests of the Amazon basin vary widely (Fearnside, 1996; Brown, 1997; Houghton, 1997; Houghton *et al*., 2001; Eva *et al*., 2003; Fearnside & Laurance, 2003), and contribute more than any other factor to the uncertainty in estimates of carbon flux from land-cover and land-use change (Houghton *et al*., 2000; Houghton, 2005). While extensive forest inventories could provide the data required for accurate determination of the sources and sinks of carbon from changes in land use, systematic on-the-ground measurements of biomass over large areas, such as the Amazon basin, are expensive and highly unlikely. Partial inventories, such as the one carried out by RADAMBRAZIL in the 1970s, and measurements at individual plots, provide information on biomass in certain forest types, but they have been insufficient for the entire region.

Houghton *et al*. (2001) compared seven methods that have been used to estimate forest biomass over the Brazilian Amazon. The methods were based on the RADAMBRAZIL inventory, on an interpolation of measurements from 44 plots, on empirical relationships between environmental factors and aboveground biomass, on percent tree cover from satellite data, and on

Correspondence: Sassan Saatchi, tel. + 818 354 1051, fax + 818 393 5285, e-mail: saatchi@congo.jpl.nasa.gov

estimates of biomass modeled with satellite-derived measurements of net primary production (NPP). Basin-wide estimates of biomass (including dead, live, and belowground) ranged over more than a factor of two, from 39 to 93 Pg C, with a mean value of 70 Pg C. Average forest biomass was 177 Mg C ha^{-1}. Data from the RADAMBRAZIL inventory produced estimates of total biomass that varied between 62.5 and 93.1 Pg C, depending on the factors used to convert stem volumes to biomass. A spatial comparison of four of the most reasonable maps showed agreement over only 5% of the Brazilian Amazon (essentially random agreement).

Estimates of biomass for the region suffer from two sources of uncertainty: (1) uncertainties associated with measurements at individual plots and (2) uncertainties in extrapolating data from individual plots to the entire basin. Measurements at individual plots are often incomplete. Full accounting requires measurement of live and dead biomass, above- and belowground biomass, lianas, palms, small trees, and other components of biomass (Brown & Lugo, 1992; Fearnside, 1992; Higuchi *et al*., 1994; Kaufman *et al*., 1998). Measurements at individual plots also suffer from other problems, such as the possible bias of data toward low (accessible) biomass, the possible bias of small plots toward large biomass, and the use of different allometric equations to calculate biomass (Brown & Lugo, 1984, 1992; Brown *et al*., 1995; Fearnside, 1997; Nelson *et al*., 1999; Houghton *et al*., 2001; Keller *et al*., 2001; Saatchi *et al*., 2007a, b see Chave *et al*., 2005 for a systematic accounting for these uncertainties and the propagation of errors).

Systematic inventory plots using statistical sampling protocols can reduce the errors in biomass estimation to 10–20% at the stand level (Brown & Lugo, 1992; Brown *et al*., 1995; Chambers *et al*., 2001; Keller *et al*., 2001). A recent study by Brown *et al*. (2000) in the Noel Kempff National Park of Bolivia showed that a statistically designed sampling technique (over 600 plots) reduced errors to $<10\%$. However, a similar sampling intensity over the Amazon basin has not been feasible because of the basin's large and inaccessible regions, and the fine-scale variation in forest structure and in species composition.

The largest uncertainty in estimating the distribution of biomass over large regions results from the techniques used in extrapolation (Houghton *et al*., 2001). In the absence of large-scale inventory plots and direct measurement of forest biomass from remote sensing data, most efforts for quantifying the distribution of biomass have focused on either interpolation techniques aimed at providing patterns of biogeographical variation of forest biomass (Malhi *et al*., 2006), or a combination of modeling and remote sensing (Potter *et al*., 2001). In this paper, we report a new method of extrapolation over the Amazon basin. By collecting data from a large number of biomass plots in a variety of forest types distributed over the basin, and by using remote sensing data sensitive to forest characteristics and environmental variables, we develop a series of metrics for extrapolating the plot data to the basin. The approach combines the strengths of both forest plots (limited in spatial coverage but providing accurate measurement of biomass) and remote sensing data (less accurate in measuring biomass directly but covering the entire region). The spatial resolution is 1 km. To cover the wide range of biomass values across the basin, we considered all vegetation types present: old growth terra firme forests, floodplains, woody and herbaceous savanna, and small forest patches along the eastern Andes and Atlantic coast. We also included the most recent land-cover map of the region (1 km resolution) in order to separate undisturbed vegetation from the ecosystems modified by human activities (secondary and degraded forests). The region of study includes all vegetation types in South America between 14°N and 20°S latitude.

The paper is divided into several sections. The first section describes the biomass plots and the remote sensing data used in this study. The section on methodology describes the approach for extrapolating the plot data over the basin. The last section presents the biomass distribution at 1 km resolution, estimates the accuracy from cross-validation, and discusses sources of errors and uncertainties.

Data

Biomass plots

Measurements of forest biomass across the Amazon basin have increased in recent years. Although these measurements do not follow a systematic inventory protocol (they vary in plot size, sampling scheme, allometric equations, and the number of components measured), they have produced the largest data set on woody biomass throughout the basin. We refer to these measurements as biomass plots and not forest inventory plots. The plots included in this study met the following criteria: (1) almost all biomass measurements were made after 1990, (2) all secondary forests plots included the years since disturbance, (3) plots were representative of larger areas. All plots, except a few in secondary forests, were sampled within a larger forest patch and, thus, could be integrated with the remote sensing data, and (4) Plots were geolocated. We located all plots on high-resolution (30 m) Landsat ETM (Enhanced Thematic Mapper) imagery acquired in late 1990s and early 2000s and, if necessary, we modified the

geographical locations to make sure they fell in the described vegetation type.

All plots contained information on aboveground live biomass (AGLB), and only occasionally included other components of forest biomass, such as dead and belowground biomass, structural information, such as the basal area and height, and average wood density. Because the most consistent quantity provided for each plot was AGLB, our analyses concentrated on the distribution of this quantity. However, the relationships between AGLB and other components of total biomass from the published literature were used to calculate total biomass over the basin (Brown & Lugo, 1992; Cairns *et al*., 1997; Delaney *et al*., 1997; Houghton *et al*., 2001). For example, aboveground dead biomass (AGDB) and the belowground biomass (BGB) in South American forests averaged 9% and 21% of AGLB, respectively (Houghton *et al*., 2001). Cairns *et al*. (1997) also showed a direct relationship between AGLB and BGB based on 85 studies of forest plots around the world.

In this study, we identified and collected data from 544 biomass plots sampled in different vegetation types throughout the basin (Fig. 1). Most of the data were not published but were contributed to this study by individual investigators. The data (plot size, vegetation cover, geographical region, date, and name of the principal investigator) are summarized in Table 1.

The biomass plots included 216 terra firme old-growth forests, 191 secondary forests of different ages, 59 woodland savannas, 40 floodplain forests, and more than 38 submontane and montane tropical forests (Table 2).

Remote sensing data

We compiled a set of remote sensing data and products from different earth observing sensors to derive metrics sensitive to the structural attributes of vegetation, landscape, and environmental variables (Table 3). The data set included both optical and microwave remote sensing sensors. Optical data were derived from 4 years of MODIS 32-day composite products. Images of the normalized difference vegetation index (NDVI) at 1 km resolution were from Huete *et al*. (2002), and leaf area index (LAI) data were from Myneni *et al*. (2002). We processed the data sets to create, first, a 1-year monthly maximum NDVI composite and average LAI images. This step improved the quality of the data by filtering out cloud cover and reducing noise in the LAI data. Then, we computed four metrics of LAI and NDVI from these data: maximum, annual mean, mean of four driest

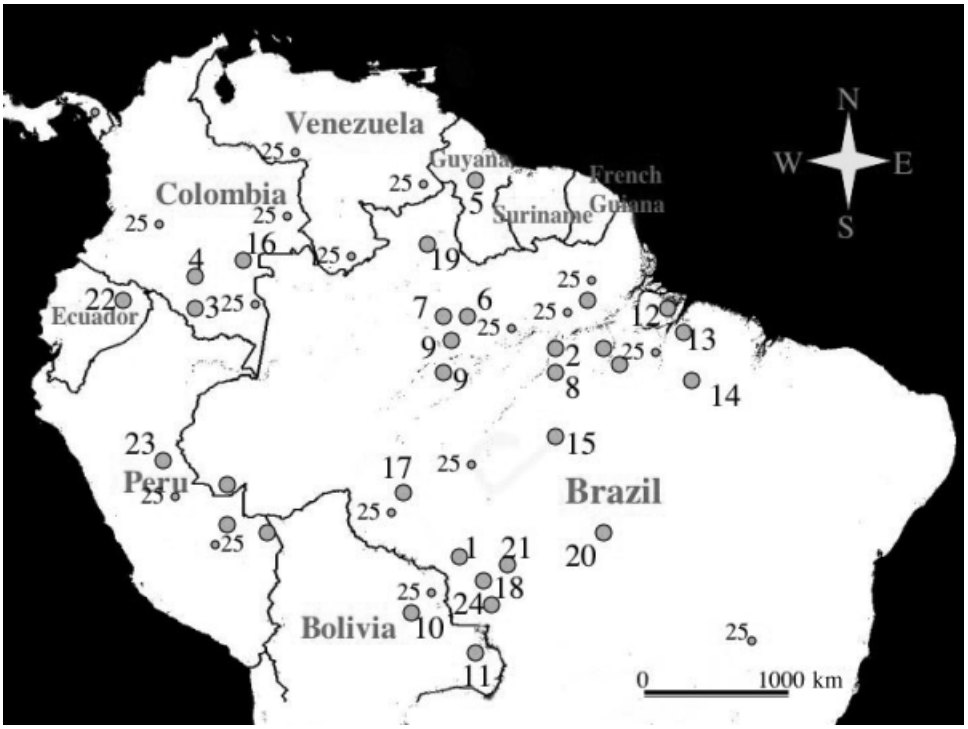

Fig. 1 Location of forest plots in the Amazon basin. Each location on the map represents several plots. Location 25 represents biomass plots scattered across the basin.

Table 1 List of biomass plot data used in this study with general locations, number of plots, vegetation types, and sources

	Reference	Location	Vegetation type	No. of plots/size
1	Cummings *et al.* (2002)	Rondonia, Brazil	Terra firme open and ecotonal forests	20 plots (0.79 ha)
2	Rice *et al.* (2004)	Tapajos, Para, Brazil	Terraa firme closed canopy dense forest	Four transects (5 ha)
3	Hoekman & Quiñones (2002)	Guaviare, Colombia	Terra firme primary and secondary forests	23 plots (0.1 ha)
4	Hoekman & Quiñones (2000)	Araracuara, Colombia	Terra firme and inundated forests	23 plots (0.1 ha)
5	van der Sanden (1997)	Mabura Hill, Guyana	Moist tropical forests	28 plots (1 ha)
6	Laurance (2002) Nascimento & Laurance (2002)	Amazonas, Brazil	Terra firme dense and fragmeneted forests	65 plots (1 ha)
7	Lucas *et al*. (2002)	Manaus, Amazonas, Brazil	Secondary and Primary forests	22 plots (0.1 ha)
8	Luckman *et al.* (1998)	Tapajos, Para, Brazil	Secondary and primary forests	18 plots (0.1 ha)
9	Steininger (2000)	Manaus, Januaca, Amazonas, Brazil	Secondary forests	18 plots (0.1 ha)
10	Steininger *et al.* (2001)	Santa Cruz, Bolivia	Inundated, liana, secondary, semideciduous, deciduous forests	26 plots (0.1 ha)
11	Brown *et al*. (2000)	Noel Kempff Natl. Park, Bolivia	Liana, inundated and evergreen forests	Six vegetation classes, from 625 plots
12	Moran & Brondizio (1998)	Marajo Island, Brazil	Secondary, logged, inundated forests	19 plots (0.1–1.0 ha)
13	Moran & Brondizio (1998)	Bragantina, Brazil	Secondary forest,	19 plots (0.1–1.0 ha)
14	Moran & Brandazio (2000)	Tome-Acu, Brazil	Secondary forest	12 plots (0.1–1.0 ha)
15	Moran & Brandazio (2000)	Altamira, Brazil	Secondary forest	16 plots (0.1–1.0 ha)
16	Moran & Brandazio (2000)	Yapu, Colombia	Secondary forest, agroforestry unit	Eight plots (0.1–1.0 ha)
17	Nelson (1998)	Acre, Brazil	Dense evergreen, bamboo forests	20 plots (0.5 ha)
18	Saatchi *et al.* (2007a, b)	Jaru, Rondonia, Brazil	Terra firme open forest	Five plots (5 ha)
19	Santos *et al.* (2003)	Mucajai, Roraima, Brazil	Dense, open evergreen, secondary forest, savanna	38 plots (0.1–0.25 ha)
20	Santos *et al.* (2003)	Comodoro, Mata Grosso, Brazil	Secondary forest, woodland, grass savanna	30 plots (0.1–0.25 ha)
21	Santos *et al.* (2003)	Jaru, Rondonia	Secondary, primary forests	18 plots (0.1–0.25 ha)
22	Pitman *et al*. (1999)	Yasuni, Ecuador	Terra firme and swamp forests	24 plots (0.1–1 ha)
23	Silman (2001)	Manu, Peru	Terra firme and floodplain forests	29 plots (1 ha)
24	Alves *et al.* (1997)	Rondonia, Brazil	Secondary, primary open forests	Nine plots (0.1 ha plots)
25	Houghton *et al.* (2001)	Brazil, Bolivia, Peru, Venezuela, Colombia	Primary, lowland, montane and submontane forests	44 plots (varying)
Total				544 plots

Location of biomass plots from Noel Kempff National Park in Bolivia were not available to this study and average biomass for six vegetation types in the park were used as training data set.

Table 2 Distribution of number of plots and biomass ranges for general vegetation types across the Amazon basin

Vegetation type	Number of plots	Average AGLB tons ha^{-1}	Standard deviation AGLB tons ha^{-1}
Old growth terra firme forest	216	254.8	103.2
Floodplain inundated forest	40	161.3	101.7
Secondary forest	191	52.9	47.5
Woodland savanna	59	20.1	30.2
Grass/shrub savanna	38	4.4	1.9

AGLB, aboveground live biomass.

Table 3 List of remote sensing data and metrics from MODIS LAI (Myneni *et al*., 2002), MODIS NDVI (http://glcf.umiacs.umd.edu/data/modis/), MODIS derived percentage tree cover (http://glcf.umiacs.umd.edu/data/modis/), QSCAT scatterometer (Long *et al*., 2001), JERS-1 radar (Saatchi *et al*., 2000), and SRTM data (http://www2.jpl.nasa.gov/srtm/)

Data record	Remote sensing sensor	Vegetation/landscape parameter	RS metrics at 1 km resolution
Monthly NDVI 2000–2004	MODIS	Vegetation type and seasonality	NDVI-1: maximum NDVI NDVI-2: annual mean NDVI NDVI-3: mean NDVI dry months NDVI-4: mean NDVI wet months
Monthly (2000–2004) leaf area index (LAI)	MODIS	Vegetation type, seasonality, productivity	LAI-1: maximum LAI LAI-2: annual mean LAI LAI-3: mean LAI wet months LAI-4: mean LAI dry months
Percent tree cover	MODIS	Forest cover, heterogeneity	VCF: continuous field product
Scatterometer backscatter Monthly composites at 1 km 1999–2004	QuikSCAT	Vegetation moisture leaf/wood density	QSCAT-H: mean backscatter HH QSCAT-V: mean backscatter VV QSCAT-SH: Std backscatter HH QSCAT-SV: Std backscatter VV
Radar backscatter at 100 m resolution (1995–1996)	JERS-1	Biomass, structural heterogeneity, cover type	JERS-DRY: dry season backscatter JERS-WET: wet season backscatter JERS-DT: dry season CV texture JERS-WT: wet season CV texture
Digital elevation (100 m resolution) 2000	SRTM	Surface elevation	SRTM-HGT: Mean Elevation SRTM-STD: ruggedness factor

NDVI, normalized difference vegetation index; SRTM, Shuttle Radar Topography Mission.

months (July, August, September, and October), and mean of four wettest months (December, January, February, and April). These data sets provided measures of vegetation greenness, seasonality (deciduousness), leaf properties, and heterogeneities. We also included MODIS-derived percentage tree cover data, available from the Global Land Cover Facility at the University of Maryland, as an indicator of tree cover and a possible surrogate for the spatial distribution of biomass (Hansen *et al*., 2002).

Microwave data from spaceborne radar instruments were used as a surrogate for forest structure and biomass. JERS-1 (Japanese Earth Resource Satellite) backscatter image mosaics for dry and wet seasons at 100 m resolution were aggregated to 1 km to produce mean backscatter and a coefficient of variation (texture) (ratio of standard deviation to mean backscatter for 100 pixels). Backscatter and texture at L-band (1.25 GHz) from this instrument are sensitive to forest structure and biomass at low densities of tree cover, such as open forests and woodland savannas (Saatchi *et al*., 1997, 2000; Luckman *et al*., 1998; Podest & Saatchi, 2002). Texture also provides information on vegetation roughness and crown size distribution, again, related to variations in biomass.

As part of the microwave remote sensing measurements, we included global NASA (National Aeronautics and Space Administration) Quick Scatterometer (QSCAT) data available in 3-day composites at 2.25 km resolution (Long *et al*., 2001). The 3-day data

over 4 years (2000–2004) were used to create average monthly composites at 1 km resolution and then further processed to produce four metrics that included annual mean and standard deviation of radar backscatter at both HH and VV polarizations (H, horizontal; and V, vertical). QSCAT radar measurements are at KU band (12 GHz) and are sensitive to surface roughness, moisture, leaf water content, and other seasonal attributes, such as deciduousness of vegetation. For vegetation types of low-density biomass, such as woodland and herbaceous savanna, measurements at different polarizations correlate positively with the aboveground biomass (Long *et al*., 2001). For areas with dense forest, the sensitivity of backscatter measurements to forest canopy roughness can help distinguish differences in biomass.

Finally, we included the Shuttle Radar Topography Mission (SRTM) digital elevation data, aggregated from 100 m resolution to 1 km. In addition to the mean elevation, the standard deviation elevation about this mean was included as an indicator of surface ruggedness or vegetation roughness. Overall, 19 remote sensing image layers representing vegetation and landscape features were included in this analysis (Table 3).

Climate data

A series of climate metrics was chosen to examine the relationship between biomass distribution and climate variables over the Amazon basin. The climate data were compiled from a number of data bases and are available from the WorldClim website (http://biogeo.berkeley.edu/) (Hijmans *et al*., 2004). These bioclimatic variables included 11 temperatures and eight precipitation metrics at 1 km spatial resolution (Table 4). The databases used to produce these climate metrics were obtained from the Global Historical Climatology Network (GHCN), the United Nations Food and Agricultural Organization (FAO), the world Meteorological Organization (WMO), and the International Center for Tropical Agriculture (CIAT), R-HYdronet, and additional country-based stations. The station data were interpolated to climate surfaces by using three independent variables (latitude, longitude, and elevation) and the thin plate smoothing spline technique (ANUSPLIN, Hutchinson, 1999). Elevation (from SRTM data) was incorporated to reduce statistical error (Hutchinson, 1999).

To these climate surfaces, we added two additional layers: the number of months rainfall was <100 mm and the number of months it exceeded 300 mm. These precipitation metrics indicate the length of the dry season and the seasonality of precipitation (Malhi & Wright, 2004).

Table 4 Description of long term averaged climate surfaces (Hijmans *et al*., 2004)

Bioclimate layer	Layer description and unit
BIO1	Annual mean temperature
BIO2	Mean diurnal range [mean of monthly (max temp–min temp)]
BIO3	Isothermality (P2/P7) (×100)
BIO4	Temperature seasonality (standard deviation ×100)
BIO5	Max Temperature of warmest month
BIO6	Min temperature of coldest month
BIO7	Temperature annual range (P5–P6)
BIO8	Mean temperature of wettest quarter
BIO9	Mean temperature of driest quarter
BIO10	Mean temperature of warmest quarter
BIO11	Mean temperature of coldest quarter
BIO12	Annual precipitation
BIO13	Precipitation of wettest month
BIO14	Precipitation of driest month
BIO15	Precipitation seasonality (coefficient of variation)
BIO16	Precipitation of wettest quarter
BIO17	Precipitation of driest quarter
BIO18	Precipitation of warmest quarter
BIO19	Precipitation of coldest quarter
Rain1	Number of months <100 mm
Rain2	Number of months >300 mm

Vegetation map

To improve the extrapolation of the biomass plots over the basin, we used a vegetation map recently produced from a fusion of remote sensing data (Saatchi *et al*., 2005a). Vegetation cover was divided into 16 types based on structure (tree density), phenology, and surface inundation conditions. The cover types were largely based on the vegetation classification of the RADAMBRASIL project, but included cover types found outside the Brazilian Amazon (Prance, 1979, 1989; Pires & Prance, 1985; Veloso *et al*., 1991; IBGE, 1997). The cover types included: terra firme forest [(1) dense closed forest, (2) open forest, (3) bamboo dominated forest, (4) liana or dry forest, (5) seasonal forest], savanna vegetation [(6) dense woodland, (7) open woodland, (8) park or shrubland savanna, (9) Grassland], wetlands [(10) closed forest, (11) open forest, (12) herbaceous, (13) mangrove, (14) open water], and anthropogenic vegetation [(15) secondary forest and plantation, (16) deforested cover types, including pas-

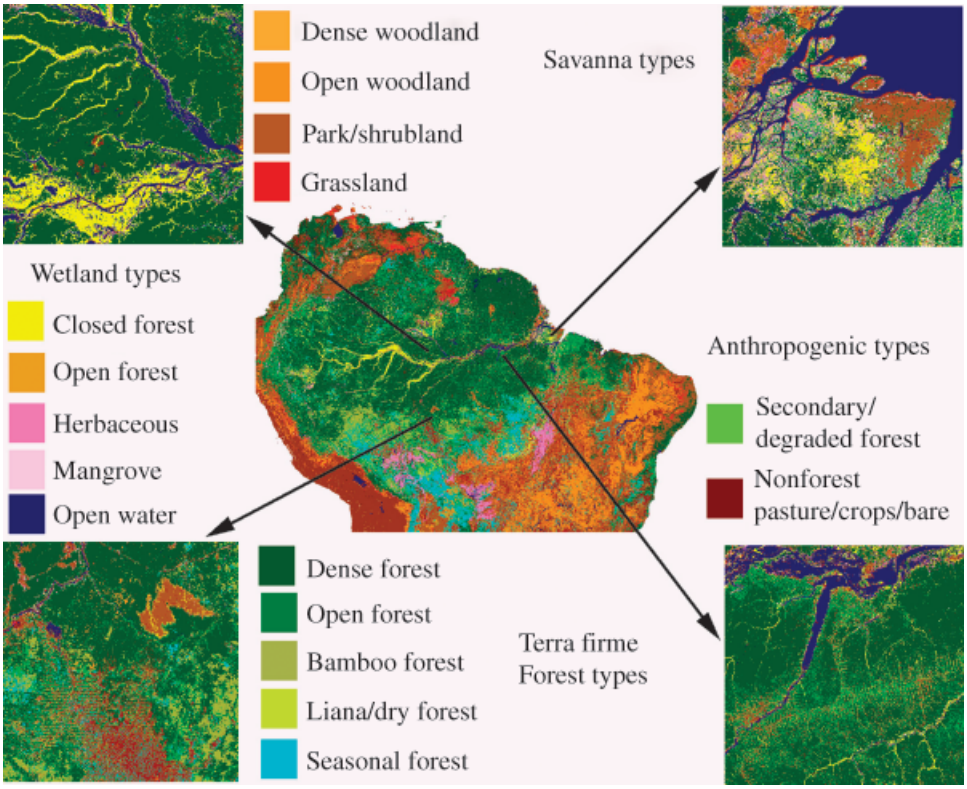

Fig. 2 Vegetation map of the Amazon basin, at 1 km spatial resolution, derived from remote sensing data (Saatchi *et al.*, 2004). The map divides the basin into 16 land cover types and open water bodies.

ture, crops, and bare areas]. The vegetation map has two advantages over other global or regional land cover maps: (1) it separates old growth forests into a number of vegetation types based on seasonality or dominance of species and (2) it separates anthropogenic cover types by using higher-resolution remote sensing data. The classification was verified with existing regional and national maps over the region. We used the map in this analysis to separate forest from nonforest types, to locate the biomass plots associated with a vegetation type (Fig. 2), and to allocate AGLB to land-cover types in the Amazon basin.

Methods

Our overall approach was to determine relationships between remote sensing metrics and AGLB from forest plots, and use these relationships directly to estimate AGLB over the entire Amazon basin. We tested several techniques, such as multivariate regression analysis and a maximum-likelihood estimator (MLE). Both methods performed poorly when tested against the plot data ($R^2 < 0.3$) because of high spatial variability of biomass at local scale and, thus, weak correlations between remote sensing metrics and biomass values. For this reason, we adopted a biomass classification approach to segment the image into different ranges of AGLB. Our methodology can be summarized in three steps: (1) classification, (2) accuracy assessment, and (3) correlation with environmental variables.

Classification

For biomass classification, we used the decision tree method (DTM) described in Simard *et al.* (2000). The method is based on the algorithm of Breiman *et al.* (1984), in which a hierarchical set of rules derived from a training data set are developed to split the input data layers into clusters associated with the class definition. DTM has been successfully applied to remote sensing data in the past because of its simplicity, efficiency, and robustness (Hansen *et al.*, 2000; Saatchi *et al.*, 2000, 2005a, 1998; Simard *et al.*, 2000, 2001). It is simple because, once the rules are determined the classification can be readily performed by using a simple program. Its efficiency is primarily due to the fact that, unlike traditional approaches (e.g. MLE), it uses only input data layers to define the classes. Finally, the methodology is robust because it does not assume any *a priori* statistical characteristics for the input data layers and, therefore,

can be applied with remote sensing data from different sensors. Moreover, the decision tree rules are explicit and allow for the identification of data layers relevant for particular class types.

We performed the classification of the AGLB in two steps. (1) We used the DTM approach to classify the AGLB of forests with biomass values above 150 Mg ha^{-1} into classes of 50 Mg ha^{-1} increments. (2) For forests with biomass values <150 Mg ha^{-1}, we developed direct regression equations from the field plots and the remote sensing data to define biomass classes of mostly 25 Mg ha^{-1} increments. The following sections summarize the methods at each step.

Forest biomass >150 Mg ha^{-1}. By concentrating on old growth terra firme and floodplain varzea forest types, we developed a training data set of 256 biomass plots (216 terra firme and 40 inundated) that included a wide range of AGLB values, the majority of which were >150 Mg ha^{-1}. The following steps summarize the overall procedure for biomass classification.

1. Using each plot's geographic coordinates, we identified the land-cover types of 256 biomass plots on the 1 km resolution vegetation map of the Amazon basin. All biomass plots were identified as belonging to either dense, open (degraded), bamboo or deciduous, dry, and floodplain or swamp forests in the classification map. If more than one plot was located on the same 1 km pixel, we used the average AGLB to represent the pixel value. These colocated plots reduced the effective number of plots from 256 to 228. The plot locations were used to create a training data set for the 19 remote sensing data layers (Table 3). We divided the training plots into seven biomass classes with 50 Mg ha^{-1} increments (i.e. 0–150, 150–200, 200–250, 250–300, 300–350, 350–400, >400.). Each class had 10–30 forest plots. For those plots that were in the middle of large contiguous forest stands and had similar landscape features (no change in elevation and land cover type), the training data were extracted from 3×3 pixels around the plot location. This approach increased the training data set to 50–100 pixels for each biomass class and allowed the development of input statistics for the DTM classifier.
2. The vegetation map was used to create a mask identifying classes of old growth and inundated forests. The mask included the dense and open terra firme forests, bamboo and liana dominated forests, seasonal, and dense and open floodplain forest cover types.
3. The DTM classifier was used to generate a biomass map with seven classes over areas masked by the vegetation map. The structure of the decision tree is determined by optimizing a cost function iteratively to assign a final node to each biomass class (Simard *et al.*, 2000). The optimization works in a global sense; it optimizes the cost function for the entire group of classes rather than individual classes. This process is performed by selecting a random sample of the training data to develop the decision tree rules and to assess the performance of the rules by predicting the biomass classes for the rest of the training pixels as an independent test data. This procedure is repeated until the highest accuracy for the independent test data is achieved. The optimized decision tree uses the most relevant data layers, and the least number of splits to obtain the classes of biomass (Simard *et al.*, 2000). We applied the decision rules from the DTM classifier to the input data layers over the region masked by the vegetation map and thereby generated a map with seven biomass classes.

Forest biomass <150 Mg ha^{-1}. To complete the mapping of biomass distribution over the Amazon basin, we included areas of woody savanna and park savanna, disturbed or secondary forests, and tree plantations. The extent and the biomass of secondary forests depend on recent rates of deforestation and land-use change. Most areas of secondary forest are small compared with the resolution of the images used in this study. However, pixels with a mixture of secondary, old growth and nonforested land were identified as anthropogenic or open forests in the vegetation map. Similarly, the woodland savanna pixels were mixtures of forest and nonforested areas. We combined these pixels from the vegetation map with the areas of the lowest biomass class (0–150 Mg ha^{-1}) to create a second vegetation mask for low biomass forests.

The estimation algorithm was developed first by extracting spectral information from the remote sensing data layers for pixels representing the forest plots with 0–150 Mg ha^{-1} of AGLB. Similar to the old growth forest case, we combined those plots that were located on the same 1 km pixel and used an average biomass to represent the pixel. After combining the pixels, we were able to create a spreadsheet of spectral information and biomass for 214 plots that included 100 secondary forests, 58 woodlands and nonforest savannas, and four mixed open forest and herbaceous swamps. The spectral data were extracted either from an individual pixel or from 3×3 pixels around the plot location wherever the biomass plot was in the middle of a contiguous large stand. We

developed an optimum multivariate regression model using a bootstrapping approach by randomly selecting half of the training data for model development and half for testing the model. The optimum model was found for the best correlation and the least error. The model was then applied to the input data layers over areas masked by the vegetation map and the resulting biomass values were classified into five classes with 25 Mg ha^{-1} increments (0–25, 25–50, 50–75, 75–100, 100–150 Mg ha^{-1}).

Accuracy assessment

Maps created from remote sensing data have at least three sources of error: (1) from the imposition of discrete classes on natural continua, (2) from the overlap of spectral characteristics of different classes, and (3) from inconsistencies between vegetation characteristics and resolution of the map. Testing and understanding the sources of errors in the biomass map are crucial in defining its applications. We used two standard approaches to assess the accuracy of the AGLB map:

1. Classification Accuracy: By classification accuracy, we refer to the probability that a randomly selected location is classified correctly on the map. Because the classification and estimation of AGLB were fully automated, we were able to embed a combination of bootstrapping and holdout procedures in the methodology to automatically estimate the expected accuracy of classification (or error of misclassification). In the holdout approach, we randomly split the training data sets into training and test subsamples. The training subsample was used to develop the decision rule or the estimation regression model, and the test subsample was used to evaluate the classification accuracy (Congalton, 1991; Steele *et al*., 1998). Assuming the training data represented the probability of biomass distribution over the basin (no *a priori* standard probability sampling technique was designed for biomass plots), the accuracy assessed from the holdout approach provides a nearly unbiased estimate of the prediction accuracy of the classification. In the bootstrap method, the holdout approach is repeated many times, by replacing the training and test data with a new random split. The individual accuracy of each classification simulation is then combined with the others to produce the expected prediction accuracy for AGLB classes (Crawford, 1989; Stehman & Czaplewski, 1998).
2. Spatial Accuracy: The above estimates of classification accuracy provide useful information about class-specific accuracy, yet there may be substantial spatial variation in accuracy across the biomass map that is not accounted for in these estimates. The spatial accuracy of the map provides the probability that any pixel or region on the map is classified correctly. Without a reference map or postclassification probability sampling data, spatial accuracy cannot be computed rigorously. Interpolation techniques, such as krigging of class accuracy over the entire map (Steele *et al*., 1998) and krigging with postclassification sampling (Kyriakidis & Dungen, 2001), are among common methods for assessing the spatial accuracy. However, these methods are time consuming and do not necessary provide reasonable assessment of accuracy in sample-deficient areas. For this analysis, we applied the bootstrap aggregation or 'bagging' method, commonly used to estimate the spatial accuracy of maps in the absence of test samples or reference maps (Brieman, 1996; Steele *et al*., 2003). By combining the DTM and the regression model, we produced one classifier from each bootstrap sampling of training data (drawn randomly with replacement) and hence one biomass map from each classifier. By repeating the bootstrap classifications many times, the probability of class membership for each pixel can be estimated as the percentage of times that pixel was classified as the optimum class. The spatial accuracy of classification is, therefore, the joint probability of the class membership for each pixel and the probability of the optimum class. We obtained estimates of probability with a Monte Carlo approximation. Note that the resulting spatial accuracy map is not an absolute assessment of biomass accuracy across basin, but a relative measure of uncertainty of the methodology across the basin.

Correlation with climate

Environmental variables such as topography, geomorphology, soil types, solar radiation, wind, temperature, and rainfall are important factors affecting the formation of the tropical forests, their diversity, structure, density, and productivity. To quantify the relationship between biomass and average bioclimatic variations over the Amazon basin, we intersected the AGLB map with the BIOCLIM variables interpolated with the digital elevation model from the SRTM data at 1km grid cells. For each AGLB class, we calculated the average and standard deviation of the climate variables and analyzed the relationship for each climate variable separately. To improve the correlations with climate variables, we increased the number of points in correlation analysis by limiting the analysis to only few biomass ranges and regrouping the biomass

classes into four levels (0–100, 100–200, 200–300, and >300).

Results

Forest biomass >150 Mg ha^{-1}

Application of the DTM to classify the AGLB of forests with biomass greater 150 Mg ha^{-1} resulted in the map shown in Fig. 3. Out of the 19 remote sensing data layers only 16 contributed to optimizing the decision rules. These layers, and the associated binary decisions are explicitly shown in Fig. 4. The DTM originally produced 30 nodes from the classification with 16 data layers. Each node was associated with one of the seven biomass classes and was used to choose the pixel class in the biomass map. However, among these nodes, only 14 were accurate enough to determine classes. These nodes were related to only 13 remote sensing layers. The rest were considered weak nodes with low accuracy in classification of pixels and were eliminated from Fig. 4. In the final map, the pixels associated with these weak nodes were corrected by applying a filter that renamed the class based on the probability of classes within a 3×3 box centered around the pixel with the low accuracy.

The interpretation of the decision tree rules for choosing biomass classes is difficult because of the large number of data layers and the multiple tree branches defining each biomass class. However, several branching rules could be considered significant in the overall classification. SRTM elevation and the ruggedness factor (standard deviation of elevation) were among the first data layers to define classes. The majority of pixels with biomass classes 4, 5, 6, and 7, referring to values above 250 Mg ha^{-1}, were located in areas below 190 m elevation and with a ruggedness of <6 m in the central Amazon. These regions had the highest probability of colocating with the geomorphological features of dense river systems. The second series of data layers important in decision rules was the QSCAT horizontal and vertical polarizations and their temporal standard deviations. These layers were sensitive to canopy roughness and moisture condition. Higher canopy roughness and moisture were, in general, associated with higher biomass values, whereas smoother canopies and lower moisture content were associated with lower biomass values. From the four NDVI metrics, the average annual NDVI was never used in the classification. The maximum and dry season NDVI metrics were primarily associated with high biomass classes (3, 4, and 5) and the wet season NDVI, with the low biomass classes (1 and 2). Among the LAI metrics, only the dry season layer was used in the final branches of the decision tree. It is possible that the dry season LAI was associated with regions of high water availability and increased leaf production during the dry season the reduced cloud cover and increased solar radiation (van Schaik *et al.*, 1993; Wright, 1996). The dry season radar backscatter and texture were also important in separating

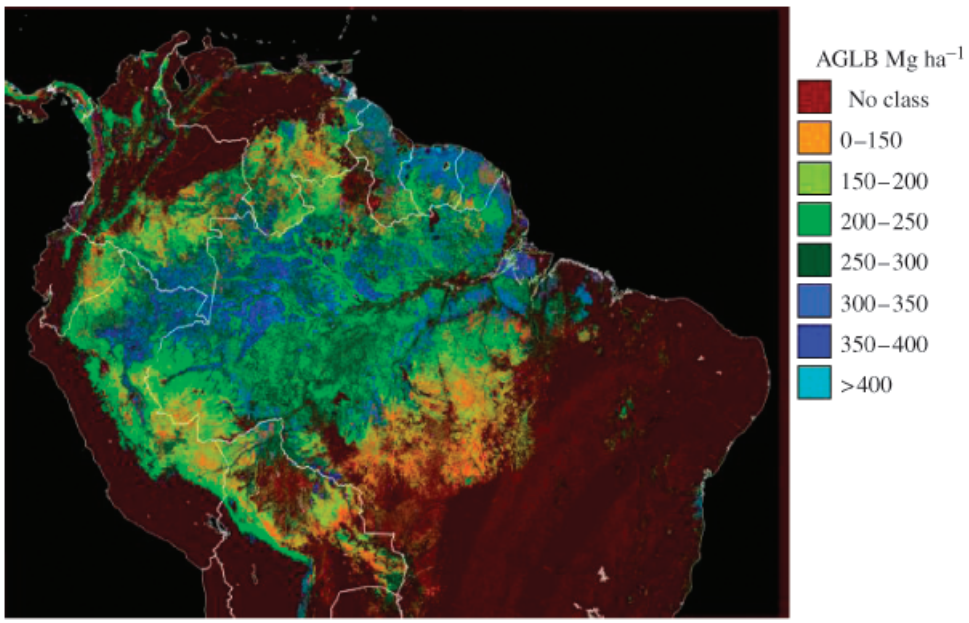

Fig. 3 Aboveground live biomass (AGLB) class map of terra firme old growth forests derived from the decision rule classifier and multiple layers of remote sensing data.

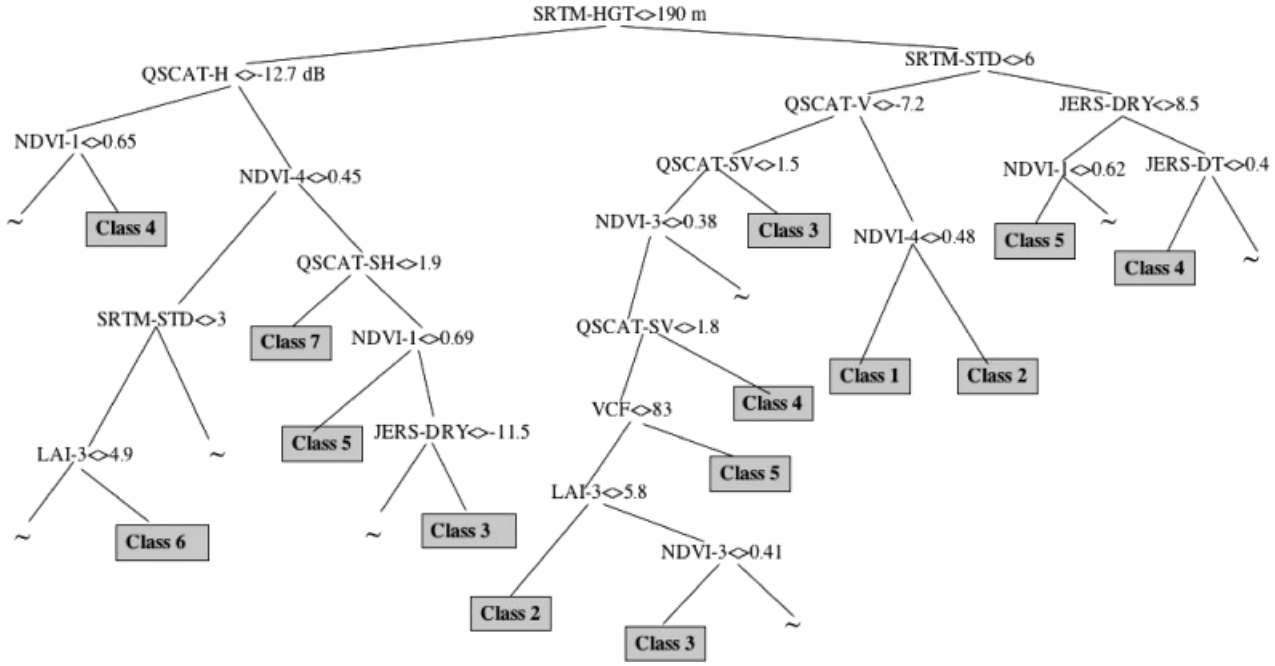

Fig. 4 Optimized decision tree rules used in the classification of the dense forest biomass map. The name of remote sensing data layers are at shown at the top of branches with their binary rules. The final nodes derived from the rules are shown at the end of each branch, with the red nodes representing the weak rules that will be finally removed by using a majority filter around the pixels associated with the weak rules.

biomass classes 3 and 4 but not for distinguishing among high biomass values.

The accuracy of the biomass classification was assessed from the combination of bootstrapping and holdout approach embedded in the DTM. The overall accuracy of classes produced from the optimum decision tree was 81% (Fig. 5), which is reasonably high given the range of AGLB chosen for the classification. The bootstrapping simulation with the holdout approach was performed 25 times to estimate the mean accuracy for each biomass class. The largest errors were for high biomass classes 6 and 7 with 73% and 58% accuracy, respectively. Given the limited sensitivity of the input data to high biomass values, such results were expected. The low and medium biomass classes had the largest number of points and the highest accuracy. We found for classes 1–5 mean prediction accuracies of 85%, 83%, 84%, 87%, and 76%, respectively. These values are considered unbiased estimates of classification accuracy. In other words, given the biomass plots and remote sensing data, the DTM performed optimally. Some of the obvious sources of errors in the classification of high biomass values were the limited number of biomass plots and the sensitivity of remote sensing data.

The results show several interesting features of AGLB distribution in the basin:

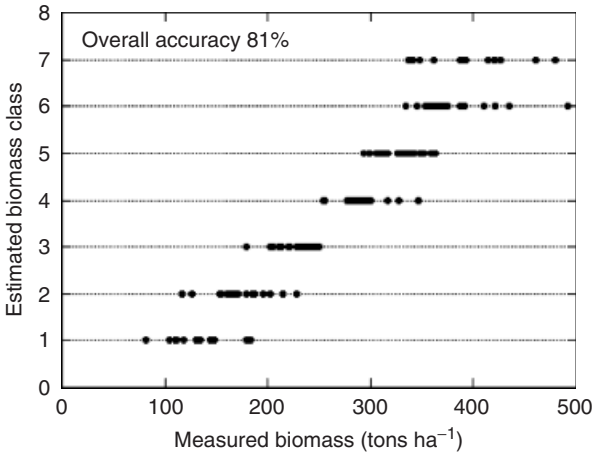

Fig. 5 Validation of biomass classification map performed internally by the decision tree method classifier using the subsample of the training pixels. The number of pixels correctly classified created an overall $R^2 = 0.81$.

1. The areas in the northeastern Amazonian region, including the Brazilian coast and the Guyanas, were estimated to have high biomass (300–400 Mg ha^{-1}). The region includes relatively intact forests because of its low human population, low agricultural potential (infertile and highly weathered forest soils),

low commercial timber volume, and inaccessibility. The climate is hot and wet and strongly influenced by the northeastern trade winds from the ocean and the intertropical convergence zone. The forest structure is multitiered with height reaching 40 m and emergent trees up to 50 m. Despite patches of savannas and low-density marsh forests around the rivers, the lowland forest in this region is expected to have high biomass values (Lindeman & Mori, 1989).

2. The central areas from west of Trombetas river to the west of Rio Negro, the region containing the main geomorphological features of the Amazon basin, with high rainfall, and elevation <100 m, were also classified as high biomass. The forests in this region are on well-drained clay or loam soil with no shortage of water. They are high in diversity, with 150–300 tree species in a single hectare and more than 500 tree species. The canopy structure is irregular, with heights ranging from 25 to 45 m, taller emergent trees and many palms. According to the biomass map, AGLB in this region ranges from 300 to 400 Mg ha^{-1} with occasional low stature forests (biomass <300 Mg ha^{-1}) on sandy soil. The main channel of the Amazon River, all major tributaries, and a large extent of varzea and igapo floodplains (biomass as high 250–300 Mg ha^{-1}) are also in this region. Along the Amazon River, the high biomass forests extend to the eastern regions of the state of Para and the Marajo Island.
3. The Western region of the Amazon basin covering a large area of the lowlands of Peru, Ecuador, Colombia, and Bolivia has biomass ranging from 200 to 300 Mg ha^{-1}. This region extends to the submontane and transitional forests near the Andean mountains and is covered by forests with open canopy, a low density of large trees, mixed with semideciduous, deciduous, bamboo, and liana trees. Data from permanent plots in this region suggest the forests are more dynamic, have a higher productivity than their counterparts in the central and eastern Amazon, and have a higher number of smaller and medium sized trees (Fearnside, 1997; Baker *et al*., 2004; Malhi *et al*., 2004).

Forest biomass <150 Mg ha^{-1}

Spectral data for forest biomass plots with <150 Mg ha^{-1} AGLB were extracted from all 19 remote sensing data layers. We found six spectral data that showed the highest correlation with the field plots (Fig. 6).

The best correlations, as expected, were based on radar backscatter measurements. In the case of JERS-1 data, the backscatter was measured at L-band (25 cm wavelength) and at 38° from nadir where the radar signal has the potential of penetrating through the forest canopy and scattering from stems. As shown in Fig. 6a, the sensitivity to biomass declines at values above 80 Mg ha^{-1} (1.9 in log scale) and almost saturates between 100 and 150 Mg ha^{-1} (2.17 in log scale). Similar results have been reported in the literature for other tropical forests (Rignot *et al*., 1997; Saatchi *et al*., 1997; Luckman *et al*., 1998).

A second significant relation was found for the annual mean of the QSCAT scatterometer measurements at both horizontal and Vertical polarizations. The QSCAT radar backscatter was measured at KU Band (2 cm wavelength) at incident angles of 46° and 54° for the H and V polarizations, respectively. At these angles, the short wavelength radar returns are highly sensitive to forest crown structure, roughness, leaf density, and moisture. According to Fig. 6b, these parameters are good surrogates for aboveground biomass in sparse woodlands and low-density forests (up to 50 Mg ha^{-1} biomass). Similar results with the spaceborne scatterometer data have been observed over savanna woodlands (Long & Hardin, 1994). Often, seasonal changes due to the deciduousness of trees or moist surface conditions may affect the scatterometer data, but these effects were not present in the annual mean backscatter data used in this study.

MODIS data, including NDVI metrics, LAI metrics, and percent tree cover, all showed reasonable correlations with the ground data. However, the best correlations with the AGLB were found for the mean NDVI of the dry season ($R^2 = 0.43$), the percent tree cover derived from the continuous field approach (DeFries *et al*., 2000) ($R^2 = 0.56$), and the LAI of dry season ($R^2 = 0.66$). The low correlation with NDVI may be related to its high sensitivity to leaf greenness, density, and seasonality. The mean NDVI for the dry season carries more information about the woody vegetation, as the grasslands and dry herbaceous understory are mostly absent during this season.

Using these six measurements (JERS-1, QSCAT H and V, dry season NDVI, percent tree cover, and dry season LAI), we developed a linear regression equation for the logarithm of AGLB in the following form:

$$\log(\mathrm{AGLB}) = 2.99 + 0.18\mathrm{LHH} + 0.0467\mathrm{QH} + 0.218\mathrm{QV} + 0.0028\mathrm{LAI} + 0.00059\mathrm{NDVI} + 0.0133\mathrm{VCF}, \quad (1)$$

where LHH is the JERS-1 radar backscatter in dB (decibels), QH and QV, respectively, represent the QSCAT H and V polarized backscatters in dB, and

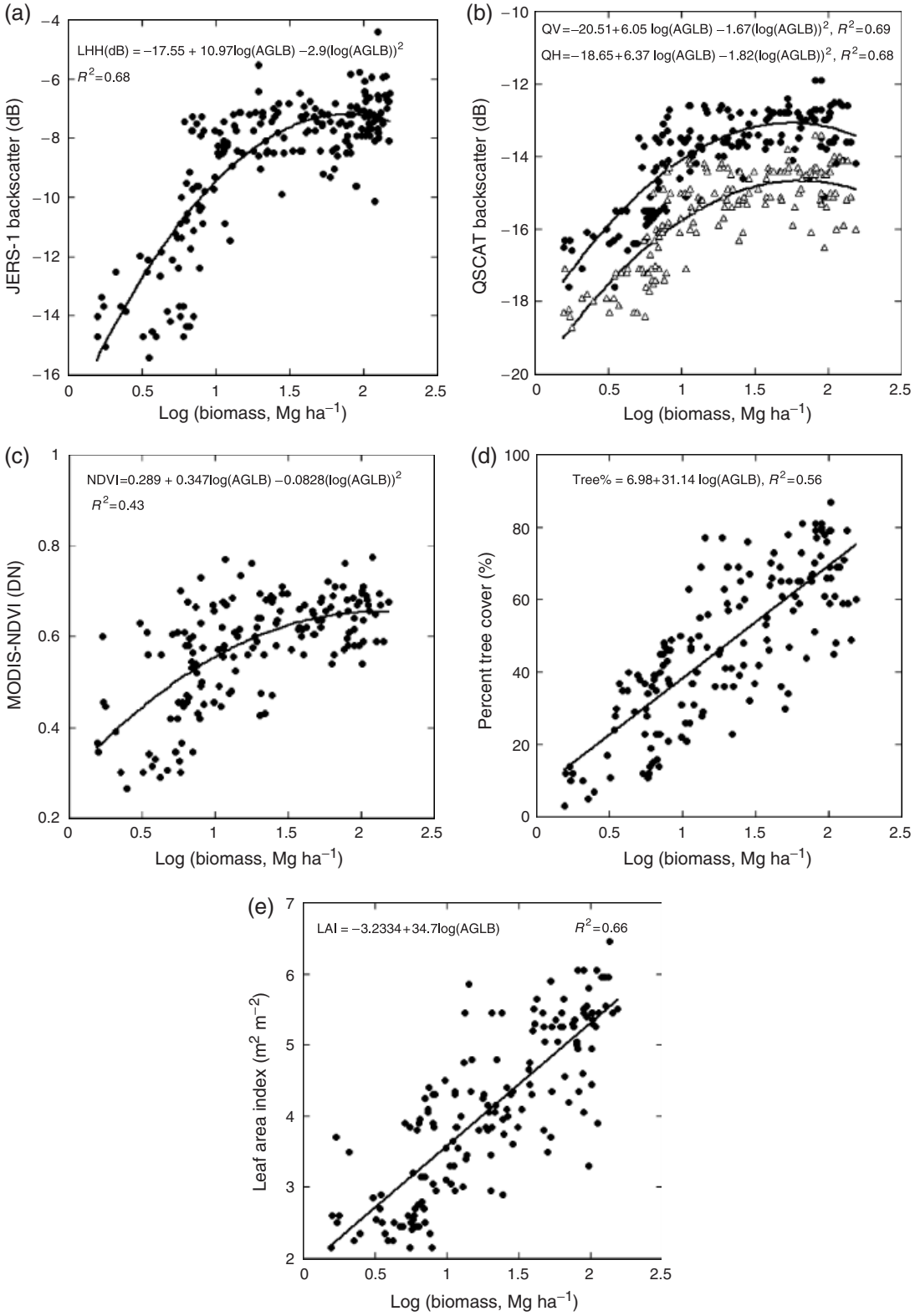

Fig. 6 Relationship between five remote sensing data and the aboveground live biomass (AGLB) of low-density forests and savanna woodlands: (a) JERS-1 radar data at L-band HH polarization with $R^2 = 0.68$, (b) QuikSCAT V- and H-polarized channels with $R^2 = 0.69$, and 0.68 respectively, (c) MODIS derived mean normalized difference vegetation index of dry season with $R^2 = 0.43$, (d) MODIS continuous field percent tree cover with $R^2 = 0.56$, and (e) MODIS derived mean LAI of dry season with $R^2 = 0.66$.

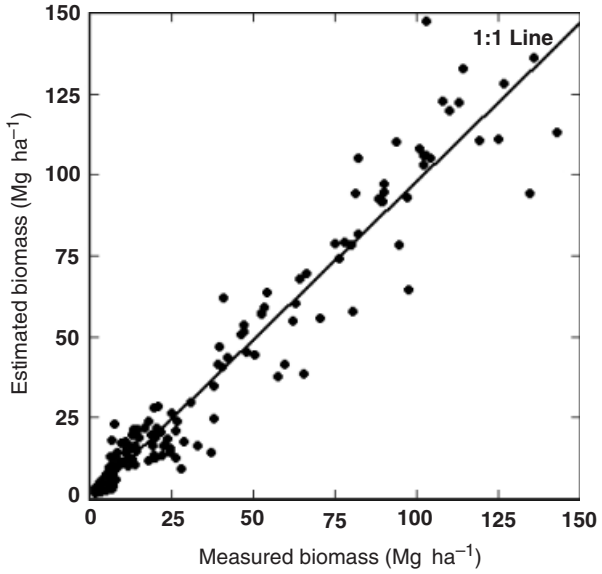

Fig. 7 Validation of biomass estimation of low density forests and savanna vegetation using the regression model derived from of combined remote sensing data ($R^2 = 0.91$, P-value < 0.0001).

VCF is the fraction of tree cover (ranging between 1 and 100). NDVI and LAI are both dry season metrics, and their values in the equation range between 0 and 1 for NDVI and between 0 and 8 for LAI. Equation (1) was developed by using an optimum set of coefficients derived from bootstrapping and holdout sampling of the training data. As discussed in Methods, for every bootstrap random sample, a linear regression model was developed and tested over the rest of the training data. Through several iterations (25) the coefficients were estimated using the best correlation and root mean square error (RMSE). This process helped select the optimum equation and guaranteed an unbiased accuracy assessment. The comparison of measured and estimated values for the optimum regression model produced $R^2 = 0.91$ and RMSE = 9.32 (Fig. 7). By partitioning the estimated biomass values into biomass classes with finer increments, the results can be represented as a map.

The overall AGLB map, combining forests with biomass $<$ and $>150\,\mathrm{Mg\,ha^{-1}}$, is shown in Fig. 8a. The spatial distribution of biomass shows that the entire area of herbaceous cover, park savanna, caatinga, and parts of open woodlands fall in the first biomass class ($0–25\,\mathrm{Mg\,ha^{-1}}$). This area includes savanna regions of the eastern and southern Amazon basin extending to the Atlantic Ocean, the savannas of Roraima in northern Brasil, the La Gran Sabana of southern Venezuela, and the areas along the Andes extending to the northwestern region of Venezuela. The majority of woodland savanna, secondary forests and regions of mixed pasture and forests fall in the second and third categories ($25–50$ and $50–75\,\mathrm{Mg\,ha^{-1}}$). The map also divides the $0–150\,\mathrm{Mg\,ha^{-1}}$ class of high-density forests into subcategories with the majority of pixels in the $100–150\,\mathrm{Mg\,ha^{-1}}$ class. The combined map includes regions outside of the Amazon basin and provides reasonable biomass values, but we cannot verify these results.

Spatial accuracy

The boostrapping aggregation or bagging approach was used to generate 25 biomass maps from combined the DTM and regression models. The spatial accuracy was computed for each pixel by multiplying the probability of the pixel being the optimum class by the classification accuracy of that class. We assumed that 25 iterations approached the Monte Carlo approximation of the bootstrap variance (Steele *et al*., 2003). However, depending on the quality and calibration of remote sensing variables at a particular pixel, 25 iterations may not be adequate to estimate the true variance. To reduce the effect of noisy or outlier pixels resulting from the classification and the estimation approach, we aggregated a map to 5 km resolution (Fig. 8b). Two general features are apparent: (1) accuracy varies with biomass. Areas with $<150\,\mathrm{Mg\,ha^{-1}}$ biomass usually have more than 80% accuracy in biomass, although the accuracy is less in some areas of old secondary forests and dense woodlands, where biomass ranges from 100 to $150\,\mathrm{Mg\,ha^{-1}}$, (2) the spatial accuracy varies within each biomass class depending on the type of vegetation or the characteristics of the remote sensing data. For example, within one biomass class, areas with higher elevation and ruggedness had relatively less accuracy than areas with flat topography. Furthermore, biomass classes in areas with high seasonal variations tend to have less accuracy than they do in areas with more stable seasonality. For example, areas dominated by bamboo forests in the southwestern Amazon, where there are seasonal variations in leaf and vegetation cover, the accuracy drops to 60%–75%.

The spatial accuracy map provides an estimate of regional uncertainty in forest biomass. It identifies where future measurements might contribute most to reducing residual uncertainties.

Discussion

The distribution of AGLB in distinct classes and the associated spatial accuracy enable us to examine the factors responsible for the magnitude and distribution of carbon stocks in the Amazon basin. We discuss three aspects of our results: (1) the relation between biomass and vegetation types, (2) comparison of the total stock

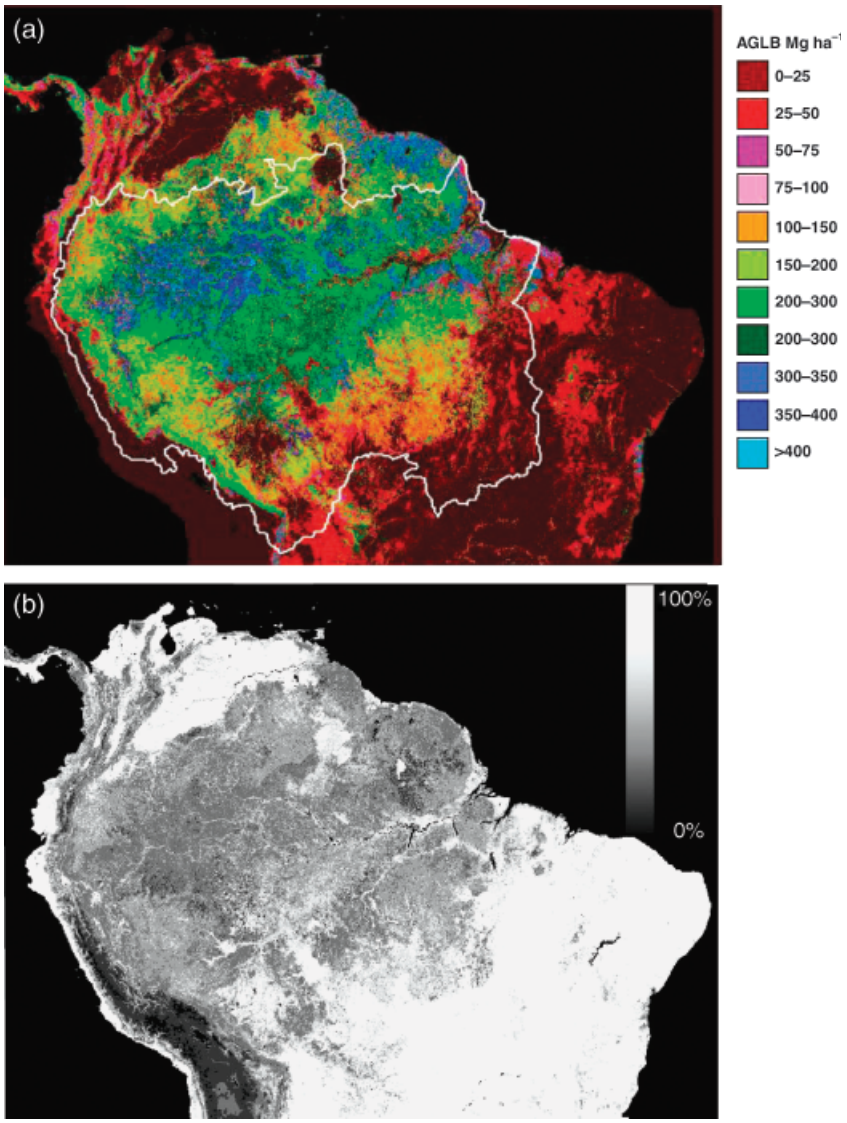

Fig. 8 Aboveground live biomass classification and the spatial accuracy assessment: (a) biomass map of the Amazon basin at 1 km spatial resolution derived from combined decision tree method and regression analysis with 11 biomass classes and overall accuracy of 88% and (b) the spatial accuracy derived from the Monte Carlo approximation of bootstrap aggregation at 5 km resolution showing the regional variations in accuracy of biomass classification.

of carbon obtained in this study with published results, and (3) the correlation of environmental variables with the patterns of AGLB.

Biomass of vegetation types

To quantify the relationship of AGLB to vegetation types of the Amazon, we intersected the vegetation map (Saatchi *et al*., 2005a) with the biomass map, and for each vegetation type we estimated the percentage of area covered by each biomass category (Table 5). The five classes of vegetation within the old growth terra firme forests (dense, open, bamboo, liana, and seasonal forests) occupy approximately 62% of the legal Amazon basin and represent the undisturbed or selectively logged forests. For each vegetation class, the biomass ranges between 150 and 350 Mg ha^{-1}. The results show that Amazonian forest biomass is extremely variable and not well correlated with vegetation type. Nor surprisingly, the techniques successful in distinguishing vegetation types need not be the ones successful for distinguishing biomass classes, and *vice versa*. The mean and the 95% confidence interval of estimated biomass for old growth forests are shown in Table 6. Vegetation types and biomass classes are to a large extent independent, and extrapolation approaches that assign ve-

Table 5 Area of the biomass classes within each general vegetation category of the Amazon basin

Biomass range Mg ha^{-1}	Old growth terra firme (62.3%) (%)	Floodplain/inundated forest (4.19%) (%)	Secondary forest (1.67%) (%)	Woodland savanna (24.47%) (%)	Grass/shrub savanna (4.79%) (%)
0–25	0	5.30	21.76	48.23	82.93
25–50	0	5.44	71.69	21.26	12.06
50–75	1.19	1.89	5.12	7.03	3.77
75–100	0.77	1.38	1.18	2.86	1.23
100–150	11.41	7.86	0.23	16.45	0
150–200	21.67	16.49	0	2.37	0
200–250	18.37	31.79	0	0.45	0
250–300	23.72	29.82	0	1.16	0
300–350	18.80	0	0	0.1	0
350–400	3.96	0	0	0	0
>400	0.66	0	0	0	0

The percent area of each vegetation type is with respect to the total area of legal Amazon (8 235 430 km^2) and the percent cover of biomass class is given with respect to the area of each vegetation class type.

Table 6 Mean and 95% confidence interval of estimated biomass densities for areas covered by terra firme and floodplain forest types

Biomass range	Dense forest	Open forest	Bamboo forest	Liana/dry forest	Seasonal/deciduous forest	Varzea flooded forest
Mean (Mg ha^{-1})	272.5	200.2	212.3	189.7	225.6	248.3
95% Confidence Interval (Mg ha^{-1})	±37.3	±59.2	±31.6	±14.5	±63.1	±23.3

getation types an average biomass value are unlikely to capture this variability. The spatial variation of the biomass is important. It is required for computing accurate estimates of carbon flux associated with deforestation and disturbance. Average values of forest biomass for regions or vegetation types may be quite different from the biomass values of the forests actually deforested.

The area of AGLB >350 Mg ha^{-1} is very small (4.5% of the area of the terra firme class). The inundated forests, including the closed and open floodplain forests and the estuary and coastal mangroves, occupy almost 4% of the basin and have lower biomass than terra firme types. The secondary forest class is approximately 1.7% of the basin and is primarily classified in the lower biomass range of 0–50 Mg ha^{-1}. The accuracy of this result cannot be independently verified from published data. However, this result is similar to estimates for the Brazilian Amazon (Alves *et al*., 1997) and implies that secondary forests are a small portion of the total biomass of the basin. Woodlands, on the other hand play a major role in total biomass distribution within the basin because they cover 24% of the basin and their biomass is 50–150 Mg ha^{-1}.

Total Amazon biomass compared with previous estimates

To estimate the total biomass of the Amazon basin and compare the results with other studies, we used published ratios of AGDB and BGB to AGLB derived from forest plots (Houghton *et al*., 2001). AGDB averaged 9% of the AGLB (range: 2–17%), and BGB averaged 21% of AGLB (range: 13–26%). We used the average ratios for all vegetation types of the basin and calculated the range of total biomass and its components for terra firme and floodplain forests in terms of carbon and the mean total biomass weighted by area. We ignored other forest types because of lack of data on the ratios for BGB and AGDB. To quantify a range of estimates, we used minimum, maximum, and the mid biomass values for each biomass class. To find the extreme ranges of the total biomass and carbon stock, we used the minimum and maximum ratios of BGB and AGDB with the minimum and maximum range of the AGLB. The results are shown in Fig. 9. The uncertainties in quantifying total biomass (TB) or the total carbon stock in the Amazon basin are mainly due to uncertainties in medium to high-range biomass classes. These class types are spatially extensive and, when used with the wide range

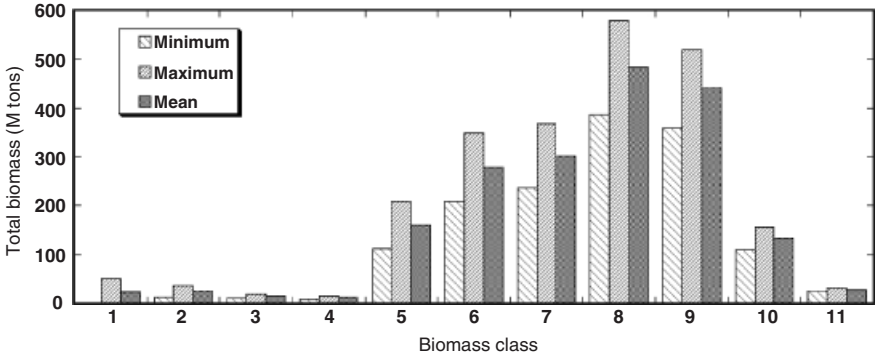

Fig. 9 Contribution of biomass classes to the total biomass of the legal Amazon basin and uncertainty calculated by using the minimum, maximum, and middle range of each class.

Table 7 Area and the magnitude of forest carbon components (AGLB, AGDB, and BGB) of terra firme and floodplain forests in the Amazon basin

Forest Type	Area (km^2)	AGLB (Pg C)	AGDB (Pg C)	BGB (Pg C)	TB (Pg C)	Mean Biomass ($Mg\,C\,ha^{-1}$)
Terra Firme	5135200	63.02	5.67	13.24	81.93	159.54
Floodplain	328825	3.25	0.29	0.68	4.22	128.33
Total	5 464 025	66.27	5.96	13.92	86.15	157.66
Range		59.19–73.34	1.32–12.27	8.62–17.23	69.13–102.84	140.81–174.49

AGLB, aboveground live biomass; ADGB, aboveground dead biomass; BGB, belowground biomass.

of ratios for AGDB and BGB, they produce large uncertainties in total biomass stock.

Converting biomass to carbon ($0.5 \times$ biomass) gives a total AGLB of 66 Pg C (range 59–73 Pg C) and a total biomass of 86 Pg C (range 69–102 Pg C) (Table 7). The latter is within the range of estimates reported by Houghton *et al*. (2001) (39–93 Pg C) and close to the high end. In contrast, mean forest biomass ($158\,Mg\,C\,ha^{-1}$ in this study) is near the lower end of the range previously reported ($100–232\,Mg\,C\,ha^{-1}$). The difference is explained, in part, by the larger area considered in this study ($5.46 \times 10^6\,km^2$, as opposed to $4 \times 10^6\,km^2$) (thus, more carbon) and, in part, by the consideration of savannas in this analysis (thus, lowering the mean biomass). The analysis by Houghton *et al*. (2001) considered only the forests of Brazilian Amazônia.

Biomass and climate variables

From the 21 BIOCLIM layers considered in this analysis, only the rainfall variables showed significant correlations with the biomass classes. We discuss several of these correlations: the mean annual rainfall, the number of months rainfall is below 100 mm, the number of months rainfall exceeds 300 mm, and the rainfall of the driest quarter. These variables relate to total water availability, the extent of the dry condition, and the magnitude and seasonality of moisture (Fig. 10). All four biomass levels are clearly separated by the number of dry months (rainfall <100 mm) (Fig. 10a). The mean evapotranspiration rate of a fully wet tropical rainforest is approximately $100\,mm\,month^{-1}$ (Salati & Marques, 1984; Shuttleworth, 1989; Malhi & Wright, 2004). When the precipitation is <100 mm per month, the forest may experience a net water deficit; hence, 100 mm threshold is a good definition for the dry season, and the number of months with precipitation below this threshold is a common definition of the length of the dry season for tropical forests. Biomass values $<100\,Mg\,ha^{-1}$ occur largely in regions with long dry season (around 6 months), while forests with $100–200\,Mg\,ha^{-1}$ occur in areas with shorter dry season (almost 4 months). The area of forests with high biomass density decrease as the number of dry months increases, indicating the consistency of moist condition for their distribution. However, there is no distinct relationship between biomass classes and the high monthly rainfalls. All four biomass categories showed similar behavior for the number of months in which rainfall exceeded 300 mm (Fig. 10b), suggesting that very high rainfall is not an important factor in controlling the biomass density. The results are similar for the relationship between mean annual rainfall and aboveground biomass. Mean annual rainfall separates only low from high biomass

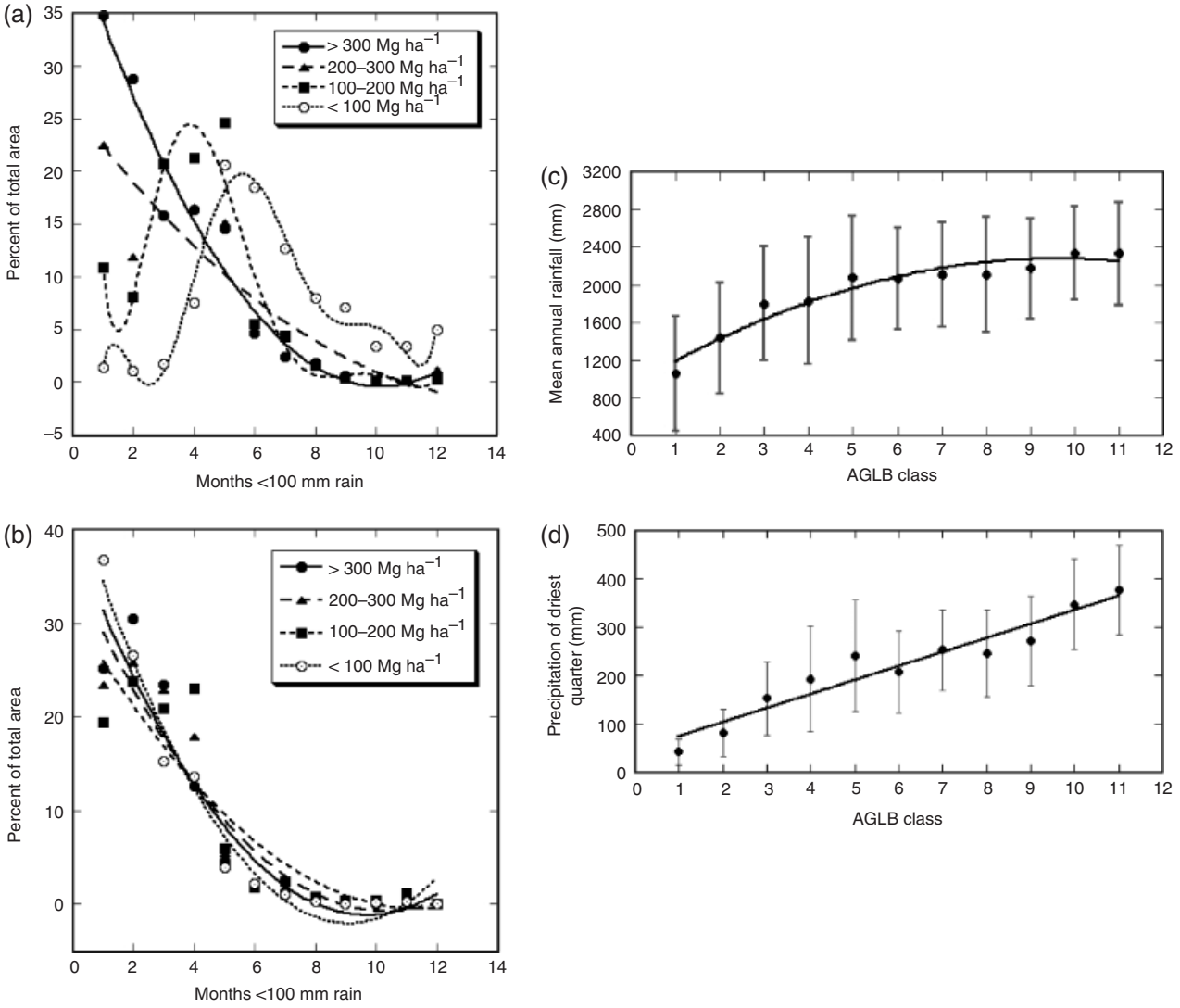

Fig. 10 Relationship between rainfall variations and the biomass distribution across the Amazon basin: (1) percent area of biomass categories falling in rainfall metric representing the number of months rainfall is <100 mm, (b) percent area of biomass falling in areas of rainfalls with number of months exceeding 300 mm, (c) the relationship between mean annual rainfall and the biomass class types, and (d) the relationship between the biomass class types and the rainfall of driest quarter.

forests and loses its sensitivity as biomass increases. The majority of high biomass forests are in regions with high rainfall distributed evenly throughout the year. However, the best relationship is found with the rainfall of the driest quarter (Fig. 10d). This finding is consistent with the observation that biomass production in tropical forests depends on the moisture available during the dry season where there is ample light and radiation (Saleska *et al*., 2003; Malhi *et al*., 2006).

Analyses with near surface air temperature show no significant correlation between temperature and biomass. In general, temperature does not vary significantly over the Amazon basin. Except at higher elevations in the Andes, where temperatures are lower, most of the basin remains between 24 and 28 °C throughout the year with annual variations of approximately 5 °C.

Conclusion

We compiled a large data set of AGLB from 544 forest plots and a large set of spatial data from remote sensing satellites to quantify the distribution of Amazonian forest biomass at fine spatial resolution. We produced a map of forest biomass classes at 1 km spatial resolution with reasonable accuracy (better than 70%) that enabled us to estimate the total carbon stock of the

basin, including the dead and belowground biomass. Our estimate of the total carbon content of the Amazon forests ranged between 77 and 95 Pg C with an average of 86 Pg C, which was within the range of published results from different approaches (Houghton *et al*., 2001). As we used the extreme ranges of dead wood and BGB ratios with low and high estimates of AGLB to compute the total biomass, the range (77–95 Pg C) must reasonably bound the total carbon stock of the basin.

Efforts to reduce this range must consider at least three questions.

1. How accurate are ground measurements of biomass over the basin? In this study, we did not address the errors associated with the aboveground biomass of forest plots. However, we know that the individual plots varied in plot size, the size of sampled trees, allometric equations, and the biomass components measured. Although a standard approach may seem desirable, it is not clear that one approach is appropriate on forest plots with different species composition and with different geographical and environmental characteristics.
2. Is it possible to reduce the uncertainty by improving the spatial resolution of data layers? This question might be tested by incorporating all available high-resolution satellite imagery and employing a multiscale approach for estimating or extrapolating biomass. One of the main sources of uncertainty in our study was the discrepancy between the resolution of images and the size of the forest plots. The spectral information obtained from 1 km resolution data is unlikely to represent the plot biomass or structure. By incorporating images at 30–100 m resolutions, we may be able to locate the plots directly on the images and remove location uncertainty, to incorporate surface heterogeneity in our calculations, and to improve the separation of the anthropogenic landscapes from forests. By using a multiscale approach, a final biomass map of 100 m resolution, or finer, might be produced, providing data sets useful in estimating the area and impact of deforestation on the carbon stock and changes in the basin.
3. What are the environmental variables responsible for the magnitude and distribution patterns of biomass density over the basin? Our results suggest a relationship between vegetation types and climate conditions. However, climatic conditions do not explain the distribution of the biomass everywhere. Soil, geomorphology, radiation and hydrological features, as well as management and land-use change, impact forest structure, species composition, and biomass. The importance of these factors can be addressed in the future as higher resolution environmental and remote sensing data layers are acquired. We expect that data available through the Large Scale Biosphere Atmosphere Experiment in Amazonia (LBA) and through other ground measurements obtained from permanent and carefully organized forest plots, such as the Amazon Forest Inventory Network (RAINFOR), will reduce the uncertainties inherent in the data used in this analysis.

Acknowledgements

We would like to thank the group of scientists, Bruce Nelson, Dirk Hoekman, Marcela Quinones, Richard Lucas, William Laurance, Marc Staining, Emilio Moran, Eduardo Brandazio, J. R. Santos, Diogenes Alves, John Terbourgh, Nigel Pitman, Miles, Silman, J. J. van der Sanden, Timothy Killeen, who shared the forest plot data with us and advised us throughout this study. We acknowledge Compton Tucker and Ranga Myneni who provided us with the optical remote sensing data. We also gratefully acknowledge the support and encouragement of Diane Wickland and Michael Keller throughout the LBA program, the LBA ecology staff for their excellent help on the use of data, and the collaboration of Brazilian organizations: Instituto Nacional de Pesquisas Espaciais (INPE) and Instituto Nacional de Pesquisas Da Amazonia (INPA). This work was performed partially at the Jet Propulsion Laboratory, California Institute of Technology, under contract from National Aeronautic and Space Administration and with the support of the LBA program. NASA's Terrestrial Ecology Program supported Houghton's participation.

References

Alves DS, Soares JV, Amaral S, Mello EMK, Almeida SAS, Da Silva OF, Silveira AM (1997) Biomass of primary and secondary vegetation in Rondônia, Western Brazilian Amazon. *Global Change Biology*, **3**, 451–461.

Baker TR, Phillips OL, Malhi Y *et al*. (2004) Variation in wood density determines spatial patterns in Amazonian forest biomass. *Global Change Biology*, **10**, 545–562.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) *Classification and Regression Trees*. Wadsworth, Pacific Grove, CA.

Brieman L (1996) Bagging predictors. *Machine Learning*, **26**, 123–140.

Brown IF, Martinelli L, Thomas W, Moreira MZ, Ferreira CAC, Victoria RA (1995) Uncertainty in the biomass of Amazonian forests: an example from Rondônia, Brazil. *Forest Ecology and Management*, **75**, 175–189.

Brown S (1997) *Estimating Biomass and Biomass Change of Tropical Forests: A Primer. FAO Forestry Paper 13*. FAO, Rome.

Brown S collected 625 biomass plots in 6 forest types as part of the Winrock International project in the Noel Kempff National park, Bolivia Information about the Winrock project and the biomass plots are given in Brown *et al*. (2000).

Brown S, Burnham M, Delaney M, Vaca R, Powell M, Moreno A (2000) *Issues and Challenges for Forest-Based Carbon-Offset Projects: a Case Study of the Noel Kempff Climate Action Project in*

Bolivia. Mitigation and Adaptation Strategies for Global Change. Springer Science Business Media BV, pp 99–121.

Brown S, Lugo AE (1992) Aboveground biomass estimates for tropical moist forests of the Brazilian Amazon. *Interciencia*, **17**, 8–18.

Cairns MA, Brown S, Helmer EH, Baumgardner GA (1997) Root biomass allocation in the world's upland forests. *Oecologia*, **111**, 1–11.

Chambers J, Santos QJ, Ribeiro RJ, Higuchi N (2001) Tree damage, allometric relationships, and above-ground net primary production in a tropical forest. *Forest Ecology and Management*, **152**, 73–84.

Chave J, Chust G, Condit R, Perez R, Lao S (2005) Error propagation and scaling for tropical forest biomass estimates. In: *Tropical Forests and Global Atmospheric Change* (eds Phillips OL, Malhi Y), pp. 155–163. Oxford University Press.

Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, **37**, 35–46.

Crawford S (1989) Extensions to the CART algorithm. *International Journal of Machine Studies*, **31**, 197–217.

Cummings DL, Boone Kauffman J, Perry DA, Flint Hughes R (2002) Aboveground biomass and structure of rainforests in the southwestern Brazilian Amazon. *Forest Ecology and Management*, **163**, 293–307.

DeFries R, Hansen M, Townshend JRG, Janetos AC, Loveland TR (2000) A new global 1 km data set of percent tree cover derived from remote sensing. *Global Change Biology*, **6**, 247–254.

Delaney M, Brown S, Lugo AE, Torres-Lezama A, Quintero NB (1997) The distribution of organic carbon in major components of forests located in five life zones of Venezuela. *Journal of Tropical Ecology*, **13**, 697–708.

Eva HD, Achard F, Stibig H-J, Mayaux P (2003) Response to comment on ''Determination of deforestation rates of the world's humid tropical forests''. *Science*, **299**, 1015b.

Fearnside PM (1992) Forest biomass in Brazilian Amazonia: comments on the estimate by Brown and Lugo. *Interciencia*, **17**, 19–27.

Fearnside PM (1996) Amazonian deforestation and global warming: carbon stocks in vegetation replacing Brazil's Amazon forest. *Forest Ecology and Management*, **80**, 21–34.

Fearnside PM (1997) Greenhouse gases from deforestation in Brazilian Amazonia: net committed emissions. *Climatic Change*, **35**, 321–360.

Fearnside PM, Laurance WF (2003) Comment on determination of deforestation rates of the world's humid tropical forests. *Science*, **299**, 1015a.

Hansen MC, DeFries R, Townshend J (2000) Global land cover classification at 1 km spatial resolution using classification tree. *International Journal of Remote Sensing*, **21**, 1331–1364.

Hansen MC, DeFries R, Townshend J, Sohlberg R, Dimiceli C, Carroll M (2002) Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment*, **83**, 303–319.

Higuchi NJ, dos Santos M, Imanaga M, Yoshida S (1994) Aboveground biomass estimate for Amazonian dense tropical moist forests. *The Memoirs of the Faculty of Agriculture, Kagoshima University*, **30**, 43–54.

Hijmans RJ, Cameron CE, Parra JL, Jones PG, Jarvis A (2004) The WorldClim interpolated global terrestrial climate surfaces. Version 1.3. Available at http://biogeo berkeley edu/

Hoekman DH, Quiñones MJ (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**, 685–696.

Hoekman DH, Quiñones MJ (2002) Biophysical forest type characterization in the Colombian Amazon by Airborne polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, **40**, 1288–1300.

Houghton RA (1997) Terrestrial carbon storage: global lessons for Amazonian research. *Ciencia e Cultura Sao Paulo*, **49**, 58–72.

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

Houghton RA, Lawrence KT, Hackler JL, Brown S (2001) The spatial distribution of forest biomass in the Brazilian Amazon: a comparison of estimates. *Global Change Biology*, **7**, 731–746.

Houghton RA, Skole DL, Nobre CA, Hackler JL, Lawrence KT, Chomentowski WH (2000) Annual fluxes of carbon from deforestation and regrowth in the Brazilian Amazon. *Nature*, **403**, 301–304.

Huete A, Didan K, Miura T, Rodriguez E, Gao X, Ferreira L (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, **83**, 195–213.

Hutchinson MF (1999) *ANUSPLIN User Guide Version 4.0 Centre for Resource and Environmental Studies*. Australian National University, Canberra.

IBGE (Instituto Brasileiro de Geografia e Estatística). (1997) *Diagnüstico Ambiental da Amazonia Legal*. CD-ROM produced by IBGE Rio de Janeiro.

Kaufman JB, Cummings DL, Hughes RF (1998) Biomass decline in Amazon forest fragments. *Science*, **282**, 1611a (in technical comments).

Keller M, Palace M, Hurtt G (2001) Biomass estimation in the Tapajos National Forest, Brazil: examination of sampling and allometric uncertainties. *Forest Ecological and Management*, **154**, 371–382.

Kyriakidis PC, Dungen JL (2001) A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions. *Environmental and Ecological Statistics*, **8**, 311–330.

Laurance WF provided 651 hectare forest biomass plots in 2002. For more information about the forest plots and the available data. see Nascimento and Laurance (2002).

Lindeman JC, Mori SA (1989) The Guianas. In: *Floristic Inventory of Tropical Countries* (eds Campbell DG, Hammond HD), pp. 375–390. New York Botanical Garden, New York, USA.

Long DG, Drinkwater M, Holt B, Saatchi S, Bertoia C (2001) Global ice and land climate studies using scatterometer image data. *EOS, Transaction of American Geophysical Union*, **82**, 1

Long DG, Hardin PJ (1994) Vegetation studies of the Amazon Basin using enhanced resolution seasat scatterometer data. *IEEE Transaction on Geoscience and Remote Sensing*, **32**, 449–460.

Lucas RM, Honzak M, Curran PJ, Foody GM, Amaral I (2002) Forest regeneration on abandoned clearances in central Amazon. *International Journal of Remote Sensing*, **23**, 965–988.

Luckman A, Baker J, Honzak M, Lucas R (1998) Tropical forest biomass density estimation using JERS-1 SAR: seasonal variation, confidence limit, and application to image mosaics. *Remote Sensing and Environment*, **63**, 126–139.

Malhi Y, Baker TR, Phillips OL *et al*. (2004) The above-ground coarse wood productivity of 104 Neotropical forest plots. *Global Change Biology*, **10**, 563–591.

Malhi Y, Wood D, Baker TR *et al*. (2006) The regional variation of aboveground live biomass in old-growth Amazonian forests. *Global Change Biology*, **12**, 1107–1138.

Malhi Y, Wright J (2004) Spatial patterns and recent trends in the climate of tropical forest regions. *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences*, **359**, 311–329.

Moran E, Brandazio E (2000) data from secondary forest plots in five sites scattered over the Amazon basin was provided to our project. For more information about the location of the plots and description, refer to Moran and Brondizio (1998).

Moran EF, Brondizio ES (1998) Land-use change after deforestation in Amazonia. In: *People and Pixels:Linking Remote Sensing and Social Science* (eds Liverman D, Mora EF, Rindfuss RR, Stern PC), pp. 94–120. National Academy Press, Washington, DC.

Myneni RB, Hoffman S, Knyazikhin Y *et al*. (2002) Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, **83**, 214–231.

Nascimento HEM, Laurance WF (2002) Total aboveground biomass in central Amazonian rainforests: a landscape-scale study. *Forest Ecology and Management*, **168**, 311–321.

Nelson BW (1998) Forest biomass inventories were conducted in Septemeber/October, 1998 in the state of Acre in the western Amazon Basin. Forest structure data were collected in 0 5 ha plots in bamboo-dominated forest (tabocal) and forests without bamboo.

Nelson BW, Mesquita R, Pereira JLG, de Souza SGA, Batista GT, Coulo LB (1999) Allometric regressions for improved estimate of secondary forest biomass in the central Amazon. *Forest Ecology and Management*, **117**, 149–167.

Pires JM, Prance GT (1985) The vegetation types of the Brazilian Amazon. In: *Key Environments: Amazonia* (eds Prance GT, Lovejoy TE), pp. 109–145. Pergamon Press, New York.

Pitman NCA, Terborgh J, Silman MR, Nunez PV (1999) Tree species distributions in an upper Amazonian forest. *Ecology*, **80**, 2651–2661.

Podest E, Saatchi S (2002) Application of multiscale texture in classifying JERS-1 Radar data over tropical vegetation. *International Journal of Remote Sensing*, **23**, 1487–1506.

Potter CS, Brooks-Genovese V, Klooster SA, Bobo M, Torregrosa A (2001) Biomass burning losses of carbon estimated from ecosystem modeling and satellite data analysis for the Brazilian Amazon region. *Atmospheric Environment*, **35**, 1773–1781.

Prance GT (1979) Notes on the vegetation of Amazonia III The terminology of Amazonian forest types subject to inundation. *Brittonia*, **31**, 26–38.

Prance GT (1989) American tropical forests. In: *Tropical Rain Forest Ecosystems: Ecosystems of the World*, **14B** (eds Lieth H, Werger MJA), pp. 99–132. Elsevier, Amsterdam, the Netherlands.

Rice AH, Pyle EH, Saleska SR *et al*. (2004) Carbon balance and vegetation dynamics in an old-growth Amazonian forest. *Ecological Applications*, **14** (Suppl.), S55–S71.

Rignot E, Salas WA, Skole DL (1997) Mapping deforestation and secondary growth in Rondonia, Brazil, using image radar and thematic mapper data. *Remote Sensing of Environment*, **59**, 167–179.

Saatchi S, Nelson B, Podest E, Holt J (2000) Mapping land cover types in the Amazon Basin using 1 km JERS-1 mosaic. *International Journal of Remote Sensing*, **21**, 1201–1234.

Saatchi S, Steinenger M, Tucker CJ, Nelson B, Simard M (2007a) Vegetation types of Amazon Basin from fusion of optical and microwave remote sensing data. *Remote Sensing of Environment*, in press.

Saatchi SS, Alvala RC, Nelson B, Yifan Y (2007b) Spatial variation in forest structure and aboveground biomass in the southwestern Brazilian Amazon. *Forest Ecology and Management* (in review).

Saatchi SS, Soares JV, Alves DS (1997) Mapping deforestation and land use in Amazon rainforest using SIR-C imagery. *Remote Sensing of Environment*, **59**, 191–202.

Salati E, Marques J (1984) Climatology of the Amazon Basin. In: *The Amazon: Limnology and Landscape Ecology of a Mighty Tropical River and Its Basin. Monographiae Biologicae*, Vol. 56 (ed. Sioli H). Dr W Junk Publishers, Dordrecht.

Saleska SR, Miller SD, Matross DM *et al*. (2003) Carbon in Amazon forests: unexpected seasonal fluxes and disturbance-induced losses. *Science*, **302**, 1554–1557.

Santos JR, Freitas CC, Araujo LS *et al*. (2003) Airborne P-band SAR applied to the above ground biomass studies in the Brazilian tropical rainforest. *Remote Sensing of Environment*, **87**, 482–493.

Shuttleworth WJ (1989) Micrometeorology of temperate and tropical forest. *Philosophical Transactions of the Royal Society of London B*, **324**, 299–334.

Silman MR (2001) Provided biomass plot data over lowland terra firme and floodplain forests of Peru, Colombia, and Ecuador in 2001. For more information refer to publications at http://www. wfu. edu/academics/biology/faculty/silman. htm.

Simard M, De Grandi F, Saatchi S, Mayaux P (2001) Mapping tropical coastal vegetation using JERS-1 and ERS-1 radar data with a decision tree classifier. *International Journal of Remote Sensing*, **23**, 1461–1474.

Simard M, Saatchi S, De Grandi GF (2000) The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 2310–2321.

Steele BM, Patterson DA, Redmond RL (2003) Toward estimation of map accuracy without a probability test sample. *Environmental and Ecological Statistics*, **10**, 333–356.

Steele MB, Winne JC, Redmond RL (1998) Estimation and mapping of misclassification probabilities for thematic land cover types. *Remote Sensing of Environment*, **66**, 192–202.

Stehman SV (1992) Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*##**58**, 1343–1350.

Stehman SV, Czaplewski RL (1998) Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment*, **64**, 331–344.

Steininger MK (2000) Satellite estimation of tropical secondary forest above-ground biomass: data from Brazil and Bolivia. *International Journal of Remote Sensing*, **21**, 1139–1157.

Steininger MK, Tucker CJ, Townshend JRG, Killeen TJ (2001) Tropical deforestation in the Bolivian Amazon. *Environmental Conservation*, **28**, 127–134.

van der Sanden JJ (1997) *Radar remote sensing to support tropical forest management Tropenbos Guyana Series 5. Tropenbos-Guyana, Georgetown, Guyana*. PhD thesis, Wageningen Agricultural University, ISBN: 90-5485-778-1.

van Schaik CP, Terborgh JW, Wright SJ (1993) The phenology of tropical forests: adaptive significance and consequences for primary consumers. *Annual Review of Ecology and Systematics*, **24**, 353–377.

Veloso HP, Rangel Filho ALR, Lima JCA (1991) *Classificâo da Vegetâo Brasileira, Adaptada a um Sistema Universal*. IBGE, Rio de Janeiro.

Wright S (1996) Phenological responses to seasonality in tropical forest plants. In: *Tropical Forest Plant Ecophysiology* (eds Stephen SM, Robin LC, Alan PS), pp. 440–460. Chapman & Hall, New York.

