Spatial Validation of the Collection 4 MODIS LAI Product in Eastern Amazonia

Luiz Eduardo O. C. Aragão, Yosio Edemir Shimabukuro, Fernando D. B. Espírito-Santo, and Mathew Williams

***Abstract*—This paper reports on the validation of the Collection 4 MODIS leaf area index (LAI) product over the Tapajós region, eastern Amazonia. The validation site is enclosed in tile h12v09 of the MODIS LAI product. The methodology to assess MODIS LAI accuracy included two main steps: 1) A multiple regression analysis for the generation of LAI surfaces, based on the relationships between field data and remote sensing information from the Enhanced Thematic Mapper Plus sensor, and between field data and topographic information from a digital elevation model; 2) the direct comparison of these LAI surfaces with the MODIS LAI surfaces. The analysis indicated that MODIS LAI is significantly overestimated for the Tapajós region by a factor of 1.18. No relationships between MODIS LAI and the validation surfaces were found. These results are indicative of a predominance of LAI retrievals by the backup algorithm, which is overcompensating LAI values at the saturation domain. The overgeneralization of the land cover layer (MOD12Q1) can be a source of uncertainties for the lookup table parameterization. Further validation efforts must be carried out over Amazonia for a quantitative quality assessment of the MODIS LAI surfaces in order to improve its accuracy.**

***Index Terms*—Amazonia, Enhanced Thematic Mapper Plus (ETM+), Large Scale Biosphere-Atmosphere Experiment in Amazonia (LBA), leaf area index (LAI), Moderate Resolution Imaging Spectroradiometer (MODIS), remote sensing.**

I. Introduction

LEAF area index (LAI), defined as one-sided green leaf area per unit ground area (square meters per square meter) in broadleaf canopies [1], is considered a key variable in scaling up forest productivity from leaf to canopy scales. The importance of climate-vegetation-carbon cycle feedbacks mean that accurate measurements of the spatial and temporal patterns of LAI are a critical requirement for ecosystem models [2]–[5]. The Earth Observing System (EOS) is, nowadays, the most effective means of collecting data and generating a range of biophysical surfaces on a regular basis (EOS Terrestrial Ecology products), including LAI. LAI surfaces have been operationally produced since June 2000 from the Moderate Resolution Imaging Spectroradiometer (MODIS) surface reflectance data [6]. The MODIS LAI product (MOD15A2) is at 1-km spatial resolution and at an eight-day interval and is publicly available through the Earth Resources Observation System (EROS) Data Center Distributed Active Archive Center.

The algorithm for the retrieval of LAI [7], [8] uses a lookup table (LUT) method (main algorithm), to achieve inversion of the three-dimensional radiative transfer (RT) problem. A backup algorithm, based on relations between the normalized difference vegetation index (NDVI) and LAI, associated with a biome classification map, is utilized to retrieve LAI values if the main algorithm fails [1]. The MODIS LAI algorithm is based on six canopy architectural types, including: 1) grasses and cereal crops; 2) shrubs; 3) broadleaf crops; 4) savannas; 5) broadleaf forests; and 6) needle leaf forests [6]. The Collection 3 MODIS Land Cover product (MOD12Q1) [9] includes a special layer containing the six biome land covers which are used to generate the Collection 4 LAI product [10], [11].

There are few published studies focusing on the evaluation of MODIS LAI product, via independent data comparison [12]–[14]. Moreover, there are contradictory results in relation to the assessment of MODIS LAI product accuracy. Whereas [12] indicated a good performance of the algorithm for African semiarid woodlands and savannas, the results of [13] indicated an overestimation for certain areas with herbaceous vegetation and poor quality backup algorithm retrievals for woody vegetation. According to [15], the anomalies found for the herbaceous vegetation were resolved in Collection 4 data reprocessing. On the other hand, the dominance of backup algorithm retrievals, in woody vegetation, was still present in the Collection 4 data. The current anomalies in Collection 4 MODIS LAI [15] and the absence of independent evaluation of this product over the Amazon basin, based on field measurements, provided the motivation in this paper for the assessment of uncertainties in MODIS LAI product.

In this study we evaluated the Collection 4 MOD15A2 product, by undertaking: 1) empirical derivation of LAI validation surfaces based on field, Landsat Enhanced Thematic Mapper Plus (ETM+) sensor, and topographic data; and 2) primary comparison between these validation surfaces and MODIS LAI surface, to quantify the uncertainties, and identify possible sources of error.

II. Methods

A. Validation Site

The study was carried out at the Tapajós region in Pará State, eastern Amazonia. This is one of the experimental sites of the Large Scale Biosphere-Atmosphere Experiment in Amazonia

Manuscript received November 2, 2004; revised July 11, 2005. Thiswork was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo under Grant Proc. 02/00985-2.

L. E. O. C. Aragão is with the Instituto Nacional de Pesquisas Espaciais (INPE), Divisão de Sensoriamento Remoto, CEP 12227-010, São José dos Campos, SP, Brazil, and also with the Centre for the Environment, University of Oxford, Oxford OX1 3QY, U.K. (e-mail: laragao@ouce.ox.ac.uk).

Y. E. Shimabukuro and F. D. B. Espírito-Santo are with the Instituto Nacional de Pesquisas Espaciais (INPE), Divisão de Sensoriamento Remoto, CEP 12227-010, São José dos Campos, SP, Brazil.

M. Williams is with the University of Edinburgh, School of GeoSciences, Edinburgh, EH9 3JN, U.K.

Digital Object Identifier 10.1109/TGRS.2005.856632

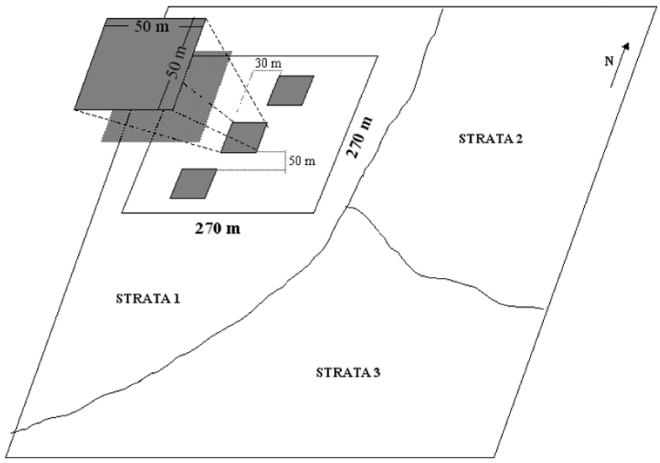

Fig. 1. Scheme of the experimental design used for field-based LAI collection and posterior combination with geospatial information.

(LBA). We selected a 13 164 km^2 area, lying between latitudes 02° 33' S and 04° 00' S and longitudes 55° 30' W and 54° 30' W. This area has a mosaic structure of secondary vegetation, pastures, and bare soil inserted in a matrix of primary forest. The primary forest has been classified into several categories; Dense Forest, with a high number of emergent tree species, Dense Forest with uniform canopy, and Open Canopy forest without palms, characterized by presence of lianas [16].

B. Sample Design and Field Measurements

The LAI measurements were taken using a stratified random sampling with a nested plot design [17]. We used the stratification procedure proposed for the validation of EOS Terrestrial Ecology products [18]. Thus, subdivisions of the global land cover classes can be obtained based on site-specific factors, such as soils and slope position, in order to reduce within-strata variation in LAI.

We stratified the study region into 23 homogeneous strata, according to vegetation, land use and topographic patterns. We sampled 13 of these strata during fieldwork in October 2002. In each strata, we marked out three 50 m × 50 m plots (total of 39) within an area of 270 m per 270 m, equivalent to 81 30-m spatial resolution EMT+ pixels (Fig. 1).

At each plot we recorded a total of 25 LAI measurements in a regular grid with 10-m spatial resolution, giving a total of 975 single measures from all 39 plots. A pair of intercalibrated LAI-2000 plant canopy analyzers (PCA) (LI-COR Inc., Lincoln, NE) was used to take simultaneous measurements within the forest and in an open area nearby. Measurements were taken in the early morning or late afternoon to minimize the incidence of direct sun on the sensor. We used a 45° view cap to avoid the computation of points with a mixture of open and closed canopy, due to the presence of tree-fall gaps. Leaf area index values for each single sample were calculated from the full five-sensor rings integration (0° to 13°, 16° to 28°, 32° to 43°, 47° to 58°, 61° to 74°).

C. Landsat EMT+ and Topographic Data Processing

We used a Landsat EMT+ image (Path 227/Row 62) acquired on July 30, 2001, during the dry season. The scene was georeferenced in the Universal Transverse Mercator (UTM) projection and in the South American Datum (SAD-69) and registered with a root mean square error (RMSE) less than one pixel (30-m spatial resolution), using topographic maps at 1 : 100 000 scale [19]. We applied the absolute radiometric correction model of [20] to each one of the six EMT+ bands (B1, B2, B3, B4, B5, and B7), to convert digital numbers to apparent reflectance. We performed a further atmospheric correction using 6S software [21]. After the preprocessing, we used the EMT+ bands (except the thermal band 6) to create a linear spectral mixing model [22], for decomposing each pixel into three components (soil, shade, and vegetation). We extracted pure spectral signatures of the components by choosing the endmembers within the image. The soil endmember was obtained on an agricultural area, which had exposed soil patches before plantation; the shade endmember was obtained on a clean water body; and the vegetation endmember was obtained on a homogeneous secondary forest canopy (Fig. 2). Solving (1), we generated the fraction images corresponding to each subpixel component.

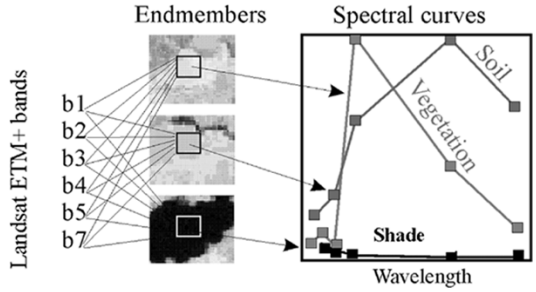

Fig. 2. Characterization of the endmembers used for the generation of the linear spectral mixing model obtained in the Landsat EMT+ image and their respective spectral curves.

$$\rho_{(i)} = a * \rho_{\text{vegetation}(i)} + b * \rho_{\text{soil}(i)} + c * \rho_{\text{shade}(i)} + e_{(i)} \quad (1)$$

where (i) is the specific band analyzed; $\rho_{(i)}$ is the pixel reflectance value; a, b, and c are the endmember proportions (vegetation, soil, and shade respectively); $\rho_{\text{vegetation}(i)}$, $\rho_{\text{soil}(i)}$, and $\rho_{\text{shade}(i)}$ are the reflectance values of each component; and $e_{(i)}$ is an error value at the specific band$_i$.

Two vegetation indexes (VIs), widely used to relate with ground-measured LAI, were also calculated, the NDVI (2) and the simple ratio (SR) (3). Both are defined in terms of red (ρ_{red}) and near-infrared (ρ_{NIR}) reflectances

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\rho_{\text{NIR}} + \rho_{\text{red}}} \quad (2)$$

$$\text{SR} = \frac{\rho_{\text{NIR}}}{\rho_{\text{red}}}. \quad (3)$$

We created a digital terrain model (DTM) from the region by the assimilation into a geographical information system (GIS) of terrain features, such as contours and drainage lines, and spot elevations, extracted from digitalized topographic maps at 1 : 100 000 scale [19]. Initially, we created a triangular irregular network (TIN) by the constrained Delaunay triangulation method, which uses the spot elevations as the triangle apex and the drainage as constraint lines [23]. Afterward, we converted the TIN, by a linear interpolation, into a regular elevation grid

(meters) with 30-m spatial resolution. We obtained the terrain slopes (degrees) via calculation of the partial derivates, for each elevation point contained in the previous elevation grid, by computing the altitude values in a 3×3 window that is moved along the grid (4)

$$\text{Slope} = \arctan \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} \tag{4}$$

where z is the elevation and x and y are the axial coordinates.

All the remote sensing and topographic data were co-registered in the same UTM (SAD-69) projection within the GIS.

D. Generation of the LAI Surfaces

1) LAI Modeling: We applied a regression approach to model LAI over the Tapajós region. Regression analysis has been extensively used to derive empirical relationship between spectral data and LAI [13], [24]–[26]. The methodology normally employed is to regress canopy biophysical properties on a simple spectral index, such as NDVI, SR or the Kauth–Thomas greenness index [27]. However, these indexes contain only a fraction of spectral information available in many image datasets and according to [13] "*there are convincing reasons for utilizing greater spectral dimensionality in regression analysis*."

In this study, we used not only a multivariate set of spectral data, including the Landsat EMT+ bands, B1 (0.45–0.52 μm), B2 (0.53–0.61 μm), B3 (0.63–0.69 μm), B4 (0.78–0.90 μm), B5 (1.55–1.75 μm), and B7 (2.09–2.35 μm), the three fraction images, soil, shade, and vegetation, and the two VIs, NDVI and SR, but also two topographic variables, the terrain elevation and slope. To evaluate the relationship between these variables, available in a grid format, and the field data, reducing possible geolocation errors, we used a routine analogous to that employed by [14]. The scale of analysis was shifted from pixel to forest stand, and then, for each stratum we calculated a mean LAI value, corresponding to the three plots. Afterward, the mean field-based LAI value of each stratum was paired with the mean value of the 81 pixels (30-m spatial resolution) in each site (Fig. 1) for the variables tested. In total, 13 samples were obtained for the spectral variables and 11 for the topographic variables. This difference was due to the absence of the topographic information for two sites in the southern part of the region.

Prior to the regression exercise, we calculated the Pearson's product-moment correlation coefficients for the relationship between LAI and the geospatial variables. We generated bivariate plots of all the potential variables and LAI to test the linearity of the relationships. Due to the dominance of the LAI values greater than three, and one single value of 1.58 (pasture site), the latter was considered an outlier, and was not considered in this analysis. Likewise, the evaluation of topographic data did not consider the disturbed sites (secondary forest and pasture). The near-linear behavior was confirmed for the tested variables.

We used a standard multiple regression technique to estimate LAI as the dependent variable of the multivariate set of independent variables. In the first five models, we input only the variables derived from the EMT+ sensor, combined two by two. The next five models assimilated both satellite and terrain data, which were combined two by two and three by three. Finally, the last model was based purely on terrain information. To avoid overfitting a given model, we imposed the rule, proposed by [26], that the number of variables to enter the model be less than one-third of the number of observations. To generate the LAI validation surfaces, we selected only one spectral model (Model 1) and one spectral-terrain model (Model 2) from the entire set of equations, based on model statistics. We create also a surface using the terrain model (Model 3).

2) Assessment of Regression Model Uncertainties: As a result of the low number of samples for model generation, no independent observations were available to test model accuracy. However, we undertook error analysis procedures to identify uncertainties in the LAI estimations derived from the three selected regression models. As an indicator of the relative robustness between the three regression models, we carried out an intercomparison considering the coefficient of determination (R^2), the standard error (SE), and the level of significance (p). We interpreted the meaning of each variable included in the equations by plotting field-based LAI for each site against the respective value of the variables. This analysis was matched with the residual distribution of the estimations among the 13 sites investigated.

3) LAI Validation Surfaces: We selected the best model to create the reference LAI validation surface, but surfaces using the other two models were also produced to support the analysis of the results. As the models were generated from the relationship between mean field LAI value and a mean of 81 pixels (270 m $\times$ 270 m), for each site and each variable, the pixel size of the overall set of spectral and topographic information was linearly converted to 270-m spatial resolution. The layers were then integrated in a GIS to calculate the LAI values for each single pixel.

E. Validation of MODIS LAI Product

The validation site is enclosed in the tile h12v09 of Collection 4 MODIS LAI. We converted the MODIS LAI product, which is originally in the Sinusoidal Projection (SIN), using the MODIS Reprojection Tool to the UTM/WGS84 projection. To overlay the MODIS and the validation surfaces (UTM/SAD-69) for direct comparison, the MODIS LAI product's Datum was converted to SAD-69, within the GIS. We used the MOD15A2 tile, containing information from the same date of the Landsat EMT+ image, July 30, 2001 (Day 209). In addition, we retrieved the surfaces from a time period immediately before (Day 201) and immediately after (Day 217) day 209, to evaluate problems due to cloud contamination on the LAI retrievals by the MOD15A2 algorithm.

To compare LAI patterns through the Tapajós, we analyzed 67 polygons distributed randomly over the entire region. We determined the mean, standard deviation and the maximum and minimum LAI values for all polygons. We then applied a multiple comparison Tukey test to determine differences between the mean values. In addition, we generated histograms of the frequency distribution of LAI values, and applied the Shapiro–Wilk's W test for normality. These tests permit the evaluation of anomalies referred to biome misclassification [28]. To investigate the spatial tendencies of the values, we regressed the MODIS LAI values on the validation surfaces

TABLE I
DESCRIPTION OF THE 13 SITES SAMPLED ALONG THE BR-163 HIGHWAY IN THE TAPAJÓS REGION. THE MEAN LAI VALUE, STANDARD DEVIATION (SD), AND COEFFICIENT OF VARIATION (CV) ASSOCIATED WITH EACH SITE WERE CALCULATED BASED ON THREE SAMPLES PER SITE

Site	Location	Latitude	Longitude	Caracterization	Mean	SD	CV
1	km 67	-02.85	-54.96	Primary Forest	**5.10**	0.97	19.02
2	km 83	-03.02	-54.97	Primary Forest	**4.61**	0.71	15.40
3	km 117	-03.35	-54.93	Pasture	**1.58**	1.01	63.92
4	km 117	-03.36	-54.95	Primary Forest	**4.83**	0.60	12.42
5	km 60	-02.83	-54.90	Primary Forest	**4.39**	0.34	7.74
6	km 88	-03.08	-54.93	Fire damaged area	**3.86**	0.45	11.66
7	km 113	-03.30	-54.94	Primary Forest	**3.49**	0.53	15.19
8	km 83	-03.05	-54.98	Primary Forest	**4.15**	0.69	16.63
9	km 84	-03.05	-54.93	Secondary Forest	**3.46**	0.50	14.45
10	km 211	-04.05	-54.94	Primary Forest	**3.73**	0.37	9.92
11	km 200	-04.01	-54.89	Primary Forest	**3.84**	0.44	11.46
12	km 184	-03.89	-54.81	Primary Forest	**3.25**	0.28	8.62
13	km 150	-03.64	-54.85	Primary Forest	**3.37**	0.59	17.51
Mean					**3.82**		
SD					**0.89**		
CV					**23.30**		

values. We evaluated the normality of the regression residuals distribution using the Shapiro–Wilk's W test and calculated the RMSE.

III. RESULTS AND DISCUSSION

A. LAI Field Measurements

We used the stratified random sampling with a nested plot design to represent the variability of LAI at the regional scale of observation. The stratification provided information on different primary forest types, and also on land use patterns across the region, such as secondary forests, pastures, and fire-damaged areas. We could describe LAI patterns in almost 60% of the landscape units over this region. Aragão *et al.* [17] have investigated the level of precision of this field-based LAI measurements across the Tapajós region. Assuming a confidence interval of 95%, Aragão *et al.* [17] found that the three 50 m × 50 m plot design was able to represent LAI variability with at least 10% and 15% level of precision at plot and unit scales, respectively.

Table I shows the mean LAI values for each of sites sampled. The mean LAI in primary forest areas varied between 5.10 ± 0.66 and 3.25 ± 0.14 in this region, and the LAI variation among woody vegetation types was 15%.

Differences between LAI values calculated by the LAI-2000 instrument (effective LAI) and the actual canopy LAI have been evaluated in the literature [29], [30]. The radiometric inversion theory (used by the LAI-2000) makes some assumptions about structural and radiative properties of plant canopies that are at odds with what is observed in real canopies. The most significant problems are: 1) foliage in the canopies is not distributed randomly, and thus foliage clumping can lead to an underestimation of LAI; 2) optical retrieval of LAI does not account for light absorption by nonphotosynthetic canopy elements, which can lead to a positive bias; and 3) radiative properties of leaves are greatly simplified, treating leaves only as absorbers, ignoring leaf transmission and scattering that are sources of negative bias [29], [30]. However, as far we know, there has been no attempt at quantification of the magnitude of these biases for tropical forest ecosystems. Thus, attempts to correct the LAI values without experimental grounds could introduce more uncertainties to the measurements. Moreover, [29] confirms that the assumption of random foliage distribution is reasonably good for natural forests.

TABLE II
CORRELATION COEFFICIENTS FOR THE RELATIONSHIP OF FIELD-BASED LAI WITH THE SPECTRAL DATA ($n = 13$) AND THE TOPOGRAPHIC DATA. ELEVATION AND SLOPE ($n = 9$). THE COEFFICIENTS OF CORRELATION MARKED IN BOLD ARE SIGNIFICANT AT 95% CONFIDENCE LEVEL

	Soil	Shade	Vegetation	NDVI	SR	B1	
LAI	**-0.74**	**0.78**	0.16	**0.63**	0.48	-0.38	
	B2	B3	B4	B5	B7	Elevation	Slope
LAI	**-0.66**	**-0.70**	**-0.67**	**-0.79**	**-0.78**	**0.75**	**-0.68**

B. Relationships Between Field-Based LAI and Geospatial Variables

The correlation analysis between LAI and the spectral and topographic variables showed significant coefficients ($p < 0.05$) for almost all the variables tested (Table II).

For EMT+ bands, the correlation coefficients ranged from 0.38 to 0.79. The strongest correlations were found for the shortwave infrared (SWIR) bands, B5 and B7, respectively, which are affected by canopy water content. All the bands were negatively correlated with the LAI. The VIs tested were less well correlated with LAI than all the individual EMT+ bands, except band 1. Similar results were achieved by [13] during the correlation analysis between EMT+ and MODIS reflectance values with LAI. These results highlight the importance of the SWIR bands for studying LAI across vegetation types.

The shade fraction was highly correlated with LAI, and the soil fraction was, as expected, negatively correlated. However, vegetation fraction did not have significant correlation with LAI. Both elevation and slope were significantly correlated with LAI. Terrain elevation was positively correlated with LAI but terrain slope was negatively correlated. [31] showed that LAI values in primary forests areas in the Tapajós are dependent on soil texture patterns, which are determined by terrain features, explaining the relationships found here.

C. LAI Modeling at the Tapajós Region

All the models could be fitted to the field LAI data within the confidence level of 95%, and with standard errors (SE) ranging from 0.35 to 0.51 (Table III). The three selected models have coefficients of determination varying from 0.76 (Model 3, $p < 0.013$) to 0.91 (Model 2, $p < 0.001$). The model that assimilated spectral and topographic data (Model 2) had the highest coefficient of determination among all the models (Table III). The multiple regression analyses indicated that shade and soil fractions, simple ratio, NDVI, and SWIR bands (B5 and B7) were the most important spectral variables for LAI assessment. However, for this region, where the majority of LAI values > 3.5, topographic information played a key role for increasing the power of Model 2 estimations.

TABLE III
REGRESSION MODELS STATISTICS INDICATING THE THREE DIFFERENT GROUPS OF MODELS, CLASSIFIED BY THE VARIABLE TYPE, AND SHOWING THE COEFFICIENT OF DETERMINATION (R^2), THE F-VALUE, THE STANDARD ERROR (SE), THE PROBABILITY (p), AND THE SAMPLE SIZE (n). IN THE EQUATIONS, "SOF" IS THE SOIL FRACTION. "SHF" IS THE SHADE FRACTION. B5 AND B7 ARE THE REFLECTANCE OF EMT+ BANDS 5 (1.55–1.75 μm) AND 7(2.09–2.35 μm), RESPECTIVELY. "SL" REPRESENTS THE TERRAIN SLOPE (DEGREES), AND "EL" IS THE TERRAIN ELEVATION (METERS)

Regression Models	R^2	F-value	Standard Error (SE)
Spectral variables			
75.15556(0.004)-0.33773*SoF(0.002)-43.76981*NDVI(0.010)	0.77	16.84	0.47
32.68637(0.008)-27.55604*NDVI(0.020)-0.448796*B7(0.002)	0.77	16.89	0.47
13.56046(0.001)-0.336373*SR(0.030)-0.164518*B5(0.001)	0.78	17.52	0.46
31.53892(0.001)-0.209621*SoF(0.001)-0.531511*SR(0.010)	0.79	19.23	0.44
12.83422(0.001)-0.406724*SR(0.010)-0.33167*B7(0.001)	0.80	19.66	0.44
Spectral+terrain variables			
-11.09153(0.004)+0.119073*ShF(0.001)-0.048701*Sl(0.030)	0.80	15.91	0.49
9.96663(0.001)-0.08213*SoF(0.002)+0.015484*El(0.015)	0.80	16.21	0.49
14.83812(0.001)-0.102229*SoF(0.001)-0.053218*Sl(0.030)	0.81	14.52	0.51
-11.95003(0.001)+16.14731*NDVI(0.001)+0.017143*El(0.002)-0.046335*Sl(0.013)	0.91	23.63	0.35
-3.300826(0.030)+0.408068*SR(0.001)+0.0189*El(0.015)-0.062382*Sl(0.017)	0.89	16.31	0.41
Terrain variables			
2.072286(0.056)+0.013912*El(0.0341)-0.03339*Sl(0.0643)	0.76	9.64	0.36

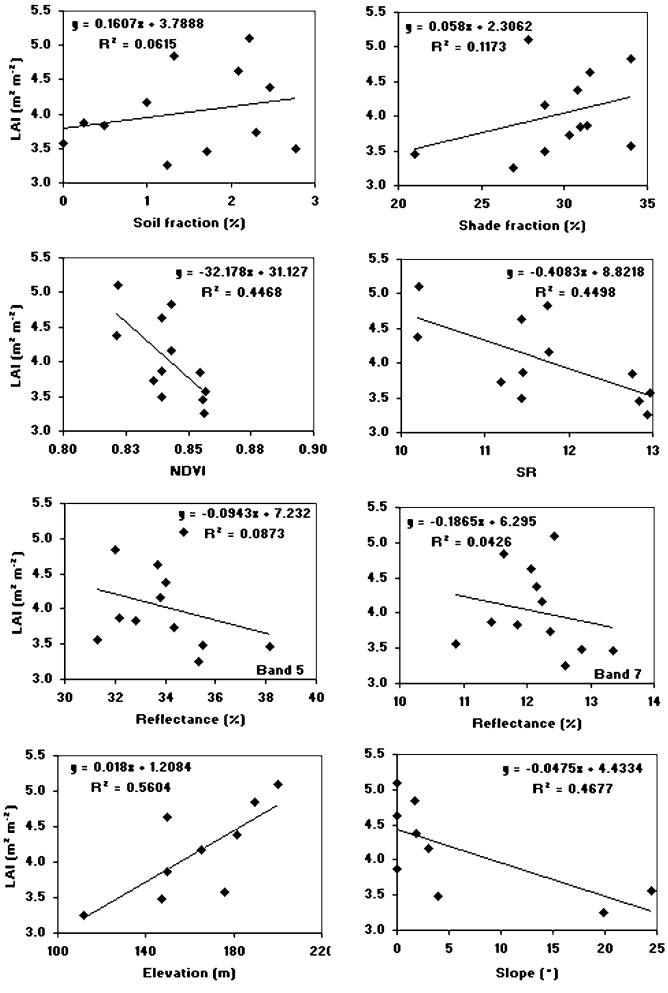

Fig. 3. Scatter plots showing the relationship between each one of the variables selected for the regression models and the field-based LAI values.

The behavior of the variables associated with the variation in LAI among sites is shown in Fig. 3. For this analysis we considered only the forest sites, excluding the sample

TABLE IV
COMPARISON AMONG THE MEAN, MINIMUM, AND MAXIMUM LAI VALUES AND THE STANDARD DEVIATION (SD) CALCULATED FOR THE VALIDATION SURFACES, THE MODIS LAI PRODUCTS, AND THE FIELD DATA. THE MEAN VALUES MARKED WITH DIFFERENT LETTERS ARE SIGNIFICANTLY DIFFERENT AT 95% CONFIDENCE LEVEL ACCORDING TO THE TUKEY TEST

	n	Mean	Minimum	Maximum	SD
MODEL 1	67	3.71[a]	0.7	4.8	0.59
MODEL 2	67	4.91[b]	1.2	6.6	0.97
MODEL 3	67	3.7[a]	1.0	5.0	0.74
MODIS (201)	67	5.11[b]	4.1	6.6	0.58
MODIS (209)	67	5.77[c]	1.6	6.6	0.71
MODIS (217)	67	6.02[c]	3.9	6.6	0.53
FIELD	13	3.82[a]	1.6	5.1	0.90

obtained in the pasture site. For woody vegetation, all the spectral variables had a negative slope for the increase of LAI, except shade and soil fractions (Fig. 3). Apart from SR, NDVI and soil fraction, the tendencies observed in Fig. 3 were corroborated by the correlation analysis (Table IV). For the vegetation index, we observed an inversion of the pattern shown in the correlation analysis (Fig. 3). This fact is explained by the influence of the pasture sample that was preserved in the correlation analysis. Signal saturation in dense canopies has obscured the expected positive relationship of NDVI and SR with LAI. This fact is probably related to an increase of shadow effects caused by emergent trees. The topographic variables showed stronger linear tendencies, in relation to LAI increases, than did the spectral variables. This fact explains the enhancement of LAI estimation by Model 2 compared to the other two models.

The residual of estimations for each one of the 13 sites varied between 0.65 and -0.50 (Fig. 4). In general, there was a systematic underestimation by the three models for the dense primary forest sites that had a LAI between 4.84 and 5.10 (sites 1, 2, and 4), and an overestimation for the open

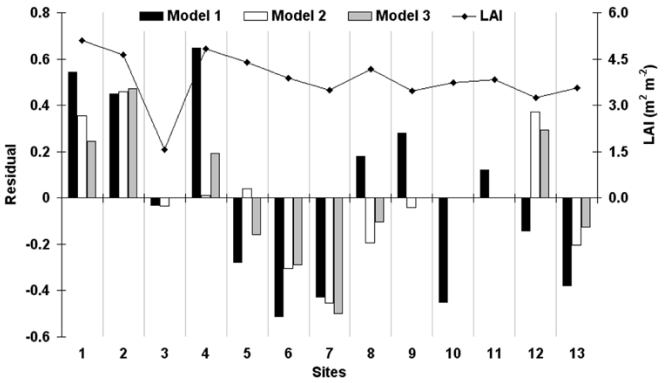

Fig. 4. Comparison of the residual of LAI estimations (bars) among the three regression models for each one of the 13 sites sampled. The solid line represents the field-based LAI values for each site.

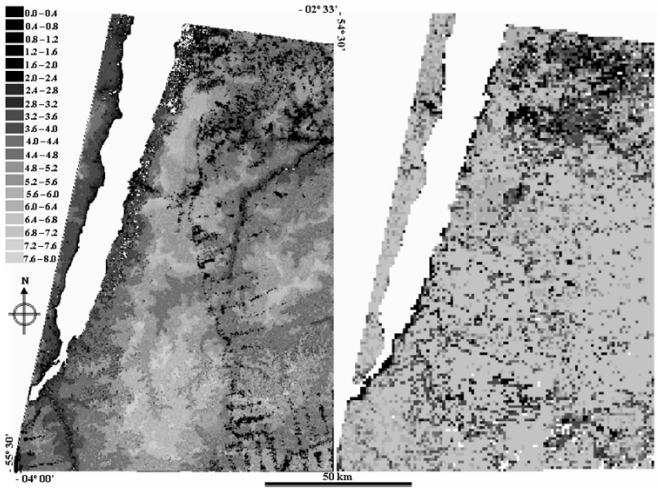

Fig. 5. Comparison between LAI surfaces generated (left) by Model 2 and (right) by the MOD15A2 algorithm for the day 209. The white areas represent pixels with no values.

canopy forests (7 and 13), the forest dominated by palms (5) and the fire-damaged forest site (6) that had the LAI ranging from 3.49 to 4.38. However, Model 2 reduced the residual of the LAI estimations in relation to Model 1 for sites 5, 6, 7, 9, and 13. The spectral model (Model 1) had the highest positive residual values at dense primary forest sites (1 and 4), indicating that the model becomes less sensitive to the variation in LAI when the values increase. This is due to a signal saturation experienced not only for the TM sensor [25] but also for the MODIS sensor [28]. Nevertheless, Model 2 reduced the error associated with the LAI estimations at the saturation domain (Fig. 4).

D. Comparisons With Collection 4 MODIS LAI Product

The comparison of the MODIS LAI surfaces with the LAI validation surface produced by Model 2 reveals two discrepancies: first, that all the three MODIS surfaces have a more homogeneous LAI pattern; and second, that MODIS LAI values are systematically higher than the estimations from Model 2 (Fig. 5). Values retrieved by the MODIS algorithm over the forest areas in the Tapajós are concentrated between LAI 6 and 7. According to [15] there is a failure of the main algorithm caused by higher values of the observed NDVI compared to the NDVI range modeled for the broadleaf forests. High NDVI values, between 0.82 and 1, lead to a constant LAI value of 6.1, reported by the backup algorithm [15]. The EMT+ data confirmed that NDVI over the forest at this site varied between 0.82 and 0.86, and the concentration of LAI values between 6 and 7 in the tile h12v09 suggest the predominance of retrievals by the backup algorithm in our study region.

The mean ($\pm$ standard deviation) LAI retrieved by MODIS (209) (5.77 ± 0.71) was significantly higher $(p < 0.05)$, at least by a factor of 1.18, than our three estimations $(3.70 \pm 0.74 - 4.91 \pm 0.97)$ and the mean of field values (3.82 ± 0.90) (Table IV). Model 2 produced a mean LAI with a standard deviation value close to that found in the field, while model 1 and 3 compressed the deviation from the mean, despite the similar mean values (Table IV). The compression of the deviation might be a consequence of the ordinary least square (OLS) regression used for generating the models. The compression of the variance in the LAI predictions from these models, using as input spectral variables, has been discussed in [26]. These authors have shown that OLS regression predicted a mean value close to that measured, but the reduction of the error is compromised by a compression of the variance of the estimated values. Nevertheless, it seems that Model 2 was not strongly influenced by this effect. The mean LAI value for the MODIS (201) LAI surface was similar to that calculated from model 2, but different $(p < 0.05)$ to the MODIS (209) and MODIS (217) LAI surfaces, this is an indication of cloud contamination over the forest area in the MODIS (201) surface. Several pixels, in forest areas, with low LAI values could be identified on this surface (not shown). Clouds that are not filtered by the algorithm generate low LAI values, due to high reflectance, consequently reducing the mean LAI.

The frequency distribution of LAI across the region showed a normal pattern for the three validation surfaces, but not for the MODIS LAI surfaces. For this analysis, we did not consider the outlier value that is related to a region with a concentration of pastures and exposed soil (Fig. 6). The distribution of the overall individual field measurements for the forested sites $(n = 900)$, excluding the measurements over pastures, also has a normal pattern [17]. The different pattern of LAI distribution found with the MODIS (201) surface, when compared with the other two MODIS surfaces, reinforces our previous conclusion about the impacts of cloudiness on MODIS LAI products. It is possible that the lack of a normal distribution for the MODIS LAI results from the overgeneralization of the land cover layer (Collection 3 MOD12Q1 product) used by the LAI algorithm. The observed displacement of LAI values to an extremely restricted range, which results in a homogenization of the MODIS LAI surface, is indicative of this problem. The LUT for the backup algorithm (LAI–NDVI relationship) is based on the corresponding mean relationships, derived from the LUT of the main algorithm [15]. If the land cover map is general for a specific region, and does not capture the structural variation of the vegetation, responsible for NDVI variation, a backup algorithm that is limited by the saturation problem will not perform well over this area.

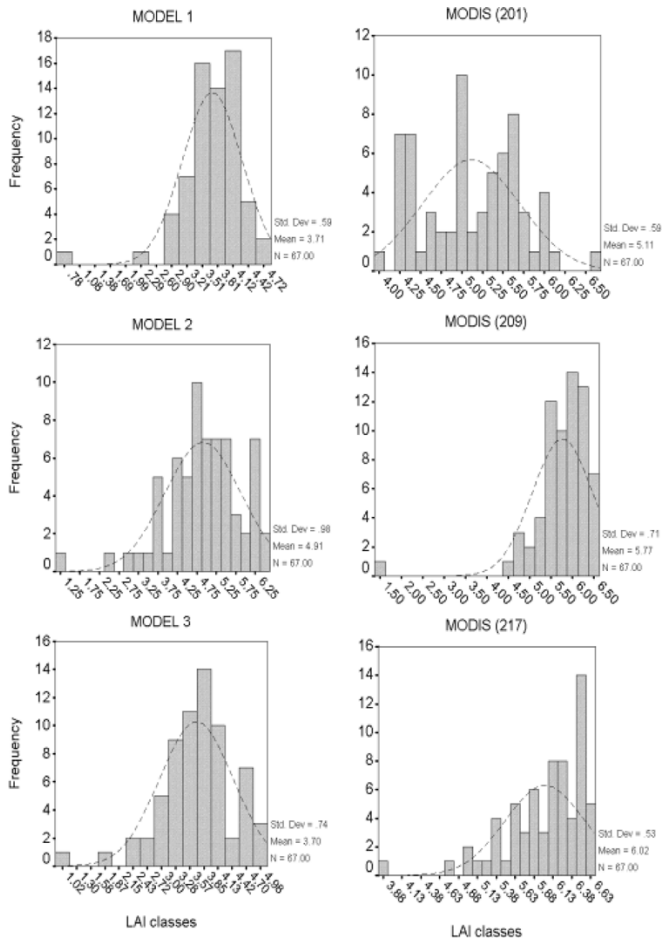

Fig. 6. Frequency distribution of the LAI values estimated by the multiple regression models and by the MOD15A2 algorithm. The bars correspond to the frequency of estimated LAI values, and the dashed lines indicate the expected normal curve.

The problem of overgeneralization of the land cover classes over the Amazon needs to be solved in order to increase the efficacy of the lookup table method used to derive MODIS LAI surfaces. An alternative is the generation of more spatially detailed maps, representing not only land cover types but also land use in the Amazon. The stratification of forest types in the Amazon can also be enhanced by the incorporation of elevation data obtained during the Shuttle Radar Topography Mission (SRTM), which generated a complete high-resolution digital topographic database.

The regression analyses of MODIS LAI on the validation surfaces did not show any tendency (Fig. 7). This result reflects the differences in the frequency distribution patterns of LAI values, found between the validation and the MODIS surfaces. The frequency distribution of the residuals of the regression between the MODIS (209) surface and the three validation models did not have a normal behavior (Fig. 8). However, for a general estimation of the regional LAI, without consideration of spatial variability, it is reasonable to apply a correction factor on the MODIS LAI based on the mean value of the residual $(\mathrm{RMSE} = 1.54\ \mathrm{m}^2 \cdot \mathrm{m}^{-2})$. This is justified by the near normal pattern found for the residuals of the regression of MODIS (209) on the Model 2 LAI surface. The RMSE value calculated in this study is close to the difference $(2\ \mathrm{m}^2 \cdot \mathrm{m}^{-2})$ between MODIS

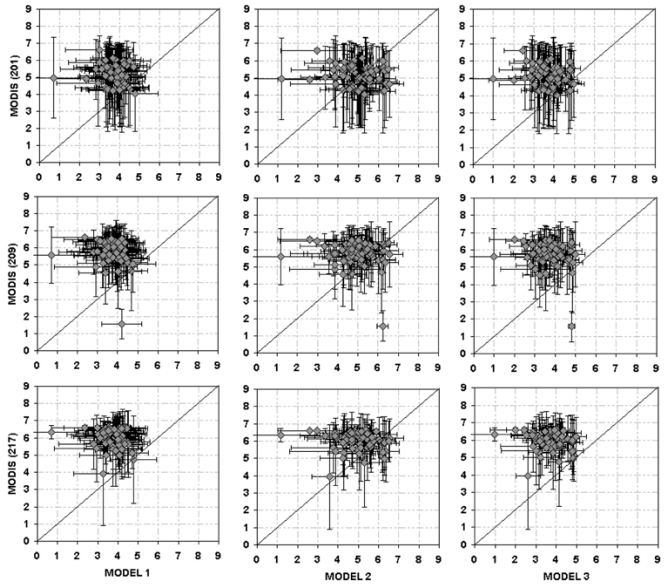

Fig. 7. Scatter plots showing the regression between the LAI retrievals by the MOD15A2 algorithm and the LAI estimated by the multiple regression models. The gray dots correspond to the mean value and the bars to the standard deviation of the pixels contained by each polygon randomly distributed on the LAI surfaces $(n = 67)$. The black line indicates the 1 : 1 relationship.

LAI and the BigFoot LAI surface for a broadleaf forest in the U.S. [13].

IV. Conclusion

The results presented here showed that the combination of spectral and terrain data could capture the spatial variability of LAI associated with the land cover and land use types in the Tapajós region. This bottom-up method is an alternative for evaluating and increasing the accuracy of MODIS LAI surfaces over the Amazon region.

The evaluated MODIS LAI surfaces overestimated LAI values and have weak spatial correlation with the LAI surface generated by the regression model in the Tapajós. However, uncertainties in the LAI values in both field-based and MODIS surfaces emphasize the need for studies focusing on LAI modeling over Amazonia. The accuracy of mapping the spatial patterns of LAI at the regional level has critical consequences for the quantification of carbon and water fluxes in Amazonia, due to error propagation during modeling of these processes. It is important to highlight that sites with mean LAI < 4, commonly found in the study region, should be carefully mapped for ecosystem modeling purposes. [32] showed that C and water fluxes estimations have a high sensitivity to changes in LAI, particularly for low values (< 4).

Cloud contamination on MODIS LAI product needs to be carefully taken into account before analyzing LAI patterns. The effect of cloudiness can lead to wrong interpretation of the results. In our study region this contamination reduced the mean LAI value retrieved by MODIS LAI algorithm and caused changes in the frequency distribution of LAI values.

The MODIS LAI product is overcompensating the saturation effect that is experienced in the MODIS LAI algorithm. The results suggest that MODIS lookup table used for LAI estimations should be improved for application in this region.

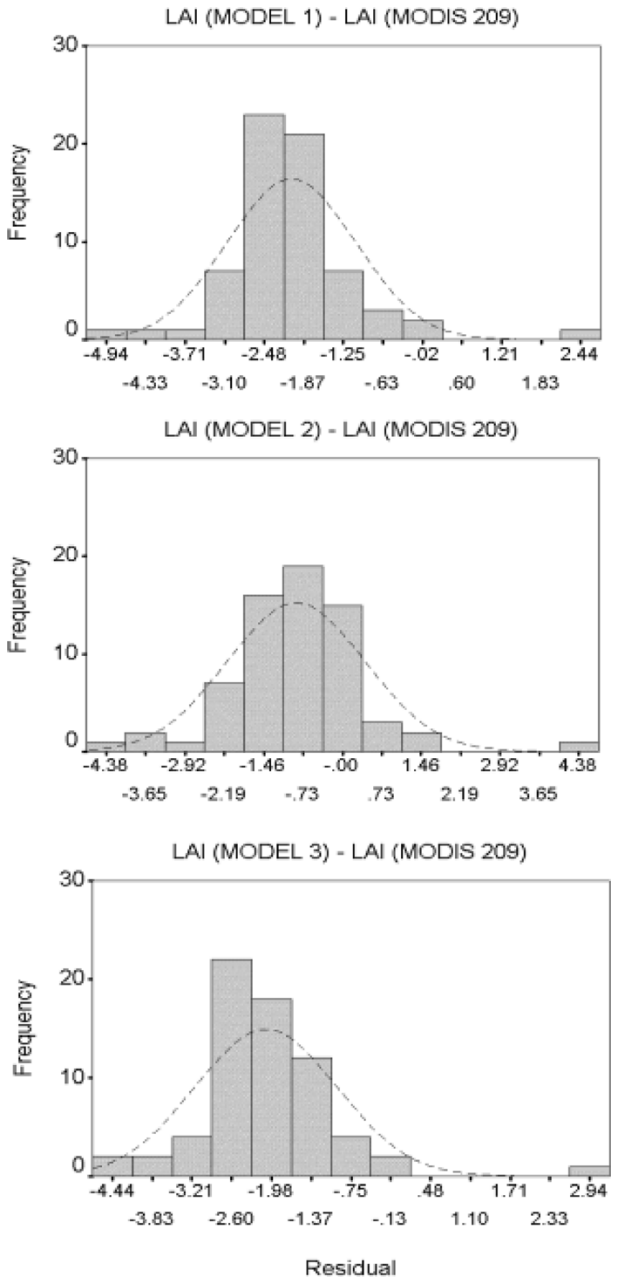

Fig. 8. Frequency distribution of the residuals obtained after the regression of LAI retrievals by the MOD15A2 algorithm on the LAI estimations by Model 2. The dashed line indicates the expected normal curve.

Further applications of this or similar methodology, systematized as a validation network across the Amazon basin, could be implemented for a quantitative assessment of the accuracy of MODIS LAI surfaces.

ACKNOWLEDGMENT

The authors thank LBA-NASA Office staff at Santarém for the logistical support and CNPq and CAPES for studentship that supported the first author. The authors would like to thank the three reviewers and Y. Malhi for their helpful comments and significant contribution to an earlier version of this paper.

REFERENCES

[1] R. B. Myneni, R. R. Nemani, and S. W. Running, "Estimation of global leaf area index and absorbed PAR using radiative trasfer models," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 6, pp. 1380–1393, Nov. 1997.

[2] S. W. Running and J. C. Coughlan, "A general model of forest ecosystem processes for regional applications I: Hydrologic balance, canopy gas exchange and primary production process," *Ecol. Model.*, vol. 42, pp. 125–154, 1988.

[3] C. S. Potter, J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney, and S. A. Klooster, "Terrestrial ecosystem production: A process model based on global satellite and surface data," *Global Biogeochem. Cycles*, vol. 7, no. 4, pp. 811–841, 1993.

[4] M. Williams, E. B. Rastetter, D. N. Fernandes, M. L. Goulden, G. R. Shaver, and L. C. Johnson, "Predicting gross primary productivity in terrestrial ecosystems," *Ecol. Appl.*, vol. 7, no. 3, pp. 882–894, 1997.

[5] H. Jiang, M. J. Apps, Y. Zhang, C. Peng, and P. M. Woodard, "Modeling the spatial pattern of net primary productivity in Chinese forests," *Ecol. Model.*, vol. 122, pp. 275–288, 1999.

[6] R. B. Myneni, S. Hoffman, Y. Knyazikhin, J. L. Privette, J. Glassy, Y. Tian, Y. Wang, S. Song, Y. Zhang, G. R. Smith, A. Lotsch, M. Friedl, J. T. Morrisette, P. Votava, R. R. Nemani, and S. W. Running, "Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data," *Remote Sens. Environ.*, vol. 83, pp. 214–231, 2002.

[7] Y. Knyazikhin, J. V. Martonchik, R. B. Myneni, D. J. Dinner, and S. W. Running, "Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data," *J. Geophys. Res.*, vol. 103, pp. 32 257–32 277, 1998.

[8] Y. Knyazikhin, J. V. Martonchik, D. J. Diner, R. B. Myneni, M. Verstraete, B. Pinty, and N. Gorbon, "Estimation of vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from atmosphere-corrected MISR data," *J. Geophys. Res.*, vol. 103, pp. 32 239–32 256, 1998.

[9] M. A. Friedl, D. K. Melver, J. C. F. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, S. Gopal, A. Schneider, and A. Cooper, "Global land cover mapping from MODIS: Algorithms and early results," *Remote Sens. Environ.*, vol. 83, pp. 287–302, 2002.

[10] B. Tan, J. Hu, P. Zhang, D. Huang, N. Shabanov, M. Weiss, Y. Knyazikhin, and R. B. Myneni, "Validation of MODIS LAI product in croplands of Alpilles, France," *J. Geophys. Res.*, 2004.

[11] D. Huang, B. Tan, W. Yang, N. V. Shabanov, Y. Knyazikhin, and R. B. Myneni, "Evaluation of collection 3 MODIS LAI product with respect to input data uncertainties—Case study for grasses," *Remote Sens. Environ.*, 2004.

[12] J. L. Privette, R. B. Myneni, Y. Knyazikhin, M. Mukelabai, G. Roberts, Y. Tian, Y. Wang, and S. G. Leblanc, "Early spatial and temporal validation of MODIS LAI product in the Southern Africa Kalahari," *Remote Sens. Environ.*, vol. 83, pp. 232–243, 2002.

[13] W. B. Cohen, T. K. Maiersperger, Z. Yang, S. T. Gower, D. P. Turner, W. D. Ritts, M. Berterretche, and S. W. Running, "Comparisons of land cover and LAI estimates derived from EMT+ and MODIS for four sites in North America: A quality assessment of 2000/2001 provisional MODIS products," *Remote Sens. Environ.*, vol. 88, pp. 233–255, 2003.

[14] Y. Wang, C. E. Woodcock, W. Buermann, P. Stenberg, P. Voipio, H. Smolander, T. Hame, Y. Tian, J. Hu, Y. Knyazikhin, and R. B. Myneni, "Evaluation of the MODIS LAI algorithm at coniferous forest site in Finland," *Remote Sens. Environ.*, vol. 91, pp. 114–127, 2004.

[15] N. Shabanov, D. Huang, W. Yang, B. Tan, Y. Knyazikhin, R. B. Myneni, D. E. Ahl, S. T. Gower, A. R. Huete, L. E. O. C. Aragão, and Y. E. Shimabukuro, "Analysis and optimization of the MODIS leaf area index algorithm retrievals over broadleaf forests," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1855–1865, Aug. 2005.

[16] Radambrasil, *Levantamento de Recursos Naturais: Folha SA.-21 Santarém, Geologia, Geomorfologia, Solos, Vegetação e uso Potencial da Terra*. Rio de Janeiro, Brazil: Projeto RADAMBRASIL/Ministerio das Minas e Energia, Dipartmento Nacional de Produção Mineral, 1976, vol. 10.

[17] L. E. O. C. Aragão, Y. E. Shimabukuro, F. D. B. Espirito-Santo, and M. Williams, "Landscape pattern and spatial variability of leaf area index in Eastern Amazonia," *Forest Ecol. Manage.*, vol. 211, no. 3, pp. 240–256, 2005.

[18] J. R. Thomlinson, P. V. Bolstad, and W. B. Cohen, "Coordinating methodologies for scaling landcover classifications from site-specific to global: Steps toward validating global map products," *Remote Sens. Environ.*, vol. 70, pp. 16–28, 1999.

[19] DSG, "Folhas SA.21-Z-B-IV, SA.21-Z-B-V, SA.21-Z-D-I, SA.21-Z-D-II, SA.21-Z-D-IV e SA.21-Z-D-V," Dir. Serviço Geográfico, Inst. asileiro de Geografia e Estatística, Brasília, Brazil, 1984.
[20] J. P. Chavez, "Image-based atmospheric corrections—Revised and improved," *Photogramm. Eng. Remote Sens.*, vol. 62, pp. 1025–1036, 1996.
[21] E. F. Vermote, D. Tanré, J. L. Deuzé, M. Herman, and J. Morcrette, "Second simulation of the satellite signal in the solar spectrum 6S: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, pp. 675–686, 1997.
[22] Y. E. Shimabukuro and J. A. Smith, "The least-squares mixing models to generate fraction images derived from remote sensing multispectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 3, pp. 16–20, May 1991.
[23] C. A. Felgueiras and M. F. Goodchild, "An incremental constrained Delaunay triangulation," Nat. Center Geograph. Inf. Anal., Santa Barbara, CA, Tech. Rep. 95-2, 1995.
[24] M. K. Butera, "A correlation and regression analysis of percent canopy closure versus TMS spectral response for selected forest site in the San Juan National Forest, Colorado," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-24, no. 1, pp. 122–129, Jan. 1986.
[25] D. P. Turner, W. B. Cohen, R. E. Kennedy, K. S. Fassnacht, and J. M. Briggs, "Relationships between leaf area index and Landsat TM spectral vegetation indexes across three temperate zone sites," *Remote Sens. Environ.*, vol. 70, pp. 52–68, 1999.
[26] W. B. Cohen, T. K. Maiersperger, S. T. Gower, and D. P. Turner, "An improved strategy for regression of biophysical variables and Landsat EMT+ data," *Remote Sens. Environ.*, vol. 84, pp. 561–571, 2003.
[27] F. G. Hall, Y. E. Shimabukuro, and K. F. Huemmrich, "Remote sensing of forest biophysical structure using mixture decomposition and geometric reflectance models," *Ecol. Appl.*, vol. 5, no. 4, pp. 993–1013, 1995.
[28] Y. Knyazikhin, J. Glassy, J. L. Privette, Y. Tian, A. Lotch, Y. Zhang, Y. Wang, J. T. Morisette, P. Votava, R. B. Myneni, R. R. Nemani, and S. W. Running, "MODIS leaf area index (LAI) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product (MOD15),", Alg. Theoretical Basis Doc., 1999. [Online]. Available: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod15.pdf..
[29] J. M. Chen and J. Cihlar, "Quantifying the effect of canopy architecture on optical measurements of leaf area index using two gap size analysis methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 3, pp. 777–787, May 1995.
[30] E. J. Hyer and S. J. Goetz, "Comparison and sensitivity analysis of instruments and radiometric methods for LAI estimation: Assessments from a boreal site," *Agricult. Forest Meteorol.*, vol. 122, pp. 157–174, 2004.
[31] L. E. O. C. Aragão, "Modeling temporal and spatial patterns of the gross primary productivity over the Tapajos region: A multiscale analysis," Ph.D. thesis (in Portuguese), Inst. Nacional Pesquisas Espaciais (INPE), Sao Jose dos Campos-SP, Brazil, 2004.
[32] M. Williams, Y. Malhi, A. D. Nobre, E. B. Rastetter, J. Grace, and M. G. Pereira, "Seasonal variation in net carbon exchange and evapotranspiration in a Brazilian rain forest: A modeling analysis," *Plant, Cell Environ.*, vol. 21, pp. 953–968, 1998.

Luiz Eduardo O. C. Aragão received the M.S. degree in environmental sciences from the University of North Fluminense, Rio de Janeiro, Brazil, and the Ph.D. degree in remote sensing from the Brazilian Institute for Space Research, São José dos Campos, Brazil, in 2000 and 2004, respectively.

Since 2004, he has been working as a Postdoctoral Research Associate in the Department of Geography and the Environment, University of Oxford, Oxford, U.K. His research is focused on the understanding of spatial and temporal patterns of carbon dynamics in the Amazon, from local to regional scales. He is working on the influence of soil fertility in carbon allocation patterns across Amazonia. He has particular interest in integrating field-based data and remote sensing for modeling the responses of forest productivity to land use and climate changes.

Yosio Edemir Shimabukuro received the B.S. degree in forestry engineering from the Rural Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, the M.S. degree in remote sensing from the Brazilian Institute for Space Research (INPE), São Paulo, and the Ph.D. degree from Colorado State University, Fort Collins, in 1972, 1977, and 1987, respectively.

From January 1992 to March 1994, he was a Visiting Scientist at the Biospheric Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD. Since 1973, he has been with INPE using satellite and ground-based remote sensing data for studying vegetation cover. He is currently a Senior Researcher at INPE. His research is focused on the monitoring and the quantification of the deforestation in the Brazilian Amazonia. He has particular interest in the use of remote sensing technologies for terrestrial ecology applications.

Fernando D. B. Espírito-Santo received the B.A. degree in forestry engineering from the Federal University of Lavras, Lavras, Brazil, and the M.Sc. degree in remote sensing from the Brazilian Institute for Space Research, São Paulo, in 2000 and 2003, respectively.

His research interest is on the spatio-temporal monitoring of land use and land cover over Amazonia using multisensor data and geostatistics techniques.

Mathew Williams received the B.A. degree in biology from the University of Oxford, Oxford, U.K., and the Ph.D. degree in environmental science from the University of East Anglia, Norwich, U.K., in 1989 and 1994, respectively.

He was a a Research Associate in the Ecosystems Center, Marine Biological Laboratory, Woods Hole, MA, between 1994 and 2000. Since 2000, he has been working as a Lecturer in the School of GeoSciences, University of Edinburgh, Edinburgh, U.K. His research interests are in operational ecological modeling, data assimilation, and vegetation–atmosphere gas exchange. He is also interested in the interaction of carbon, nutrient, and hydrological cycles, primary productivity and allocation using ecosystem models.