



MODIS NDVI time-series allow the monitoring of *Eucalyptus* plantation biomass

Guerric le Maire ^{a,b,*}, Claire Marsden ^{a,c}, Yann Nouvellon ^{a,d}, Clovis Grinand ^e, Rodrigo Hakamada ^f, José-Luiz Stape ^g, Jean-Paul Laclau ^{a,h}

^a CIRAD, UMR Eco&Sols, 2 Place Viala, 34060 Montpellier, France

^b CIRAD, UMR TETIS, Maison de la Télédétection, 34093 Montpellier Cedex 5, France

^c SupAgro, UMR Eco&Sols, 2 Place Viala, 34060 Montpellier, France

^d USP, Atmospheric Sciences Department, Rua do Matão 1226, 05508-090 São Paulo, Brazil

^e IRD, UMR Eco&Sols, 2 Place Viala, 34060 Montpellier, France

^f International Paper do Brasil, Rodovia SP 340, Km 171, 13.840-970 Mogi Guaçu, SP, Brazil

^g Department of Forestry and Environmental Sciences, North Carolina State University, Raleigh, NC 27695, United States

^h USP, Ecology Department, Rua do Matão 321, 05508-900, São Paulo, Brazil

ARTICLE INFO

Article history:

Received 23 February 2011

Received in revised form 23 May 2011

Accepted 29 May 2011

Available online 1 July 2011

Keywords:

Aboveground biomass

Moderate Resolution Imaging

Spectroradiometer

CBERS

WorldClim

Brazil

Forest

Fast-growing plantations

ABSTRACT

The use of remote sensing is necessary for monitoring forest carbon stocks at large scales. Optical remote sensing, although not the most suitable technique for the direct estimation of stand biomass, offers the advantage of providing large temporal and spatial datasets. In particular, information on canopy structure is encompassed in stand reflectance time series. This study focused on the example of *Eucalyptus* forest plantations, which have recently attracted much attention as a result of their high expansion rate in many tropical countries. Stand scale time-series of Normalized Difference Vegetation Index (NDVI) were obtained from MODIS satellite data after a procedure involving un-mixing and interpolation, on about 15,000 ha of plantations in southern Brazil. The comparison of the planting date of the current rotation (and therefore the age of the stands) estimated from these time series with real values provided by the company showed that the root mean square error was 35.5 days. Age alone explained more than 82% of stand wood volume variability and 87% of stand dominant height variability. Age variables were combined with other variables derived from the NDVI time series and simple bioclimatic data by means of linear (Stepwise) or nonlinear (Random Forest) regressions. The nonlinear regressions gave r-square values of 0.90 for volume and 0.92 for dominant height, and an accuracy of about 25 m³/ha for volume (15% of the volume average value) and about 1.6 m for dominant height (8% of the height average value). The improvement including NDVI and bioclimatic data comes from the fact that the cumulative NDVI since planting date integrates the interannual variability of leaf area index (LAI), light interception by the foliage and growth due for example to variations of seasonal water stress. The accuracy of biomass and height predictions was strongly improved by using the NDVI integrated over the two first years after planting, which are critical for stand establishment. These results open perspectives for cost-effective monitoring of biomass at large scales in intensively-managed plantation forests.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

International agreements and regional markets now give an economical value to forest carbon stocks in order to encourage countries to increase or maintain these stocks for climate mitigation purposes. Such economical assessments require precise and reliable methods to estimate forest carbon stocks at large scales, for instance through the development of remote-sensing applications, which has become an active research field (Baker et al., 2010; Goetz et al., 2009). The use of satellite-based estimations of carbon biomass is a

promising solution in terms of (i) cost- and time-effectiveness compared to large-scale field inventories, (ii) integration of forest spatial variability in regional synopses of carbon stocks, and (iii) reactivity to high-impact disturbances such as deforestation and afforestation.

Different types of imagery have been used to assess forest carbon biomass from remote sensing. Interferometric radar and lidar data are the most promising techniques for forest biomass estimation, and it is recognized that optical imagery cannot reach the same level of accuracy (Patenaude et al., 2005). However, the immediate need for biomass and biomass change estimates across long time-periods and large scales cannot be satisfied by active remote-sensing techniques, meaning that optical imagery solutions are still essential (Powell et al., 2010). The main limitation with optical remote-sensing is its rapid

* Corresponding author at: CIRAD, UMR Eco&Sols, 2 Place Viala, 34060 Montpellier, France. Tel.: +33 4 67 54 87 64; fax: +33 4 67 54 87 00.

E-mail address: guerric.le_maire@cirad.fr (G. le Maire).

saturation with forest biomass: the reflectance signal in visible and near infrared (NIR) is mostly correlated with the green leaf area index (LAI) and canopy cover of the vegetation and saturates around an LAI of 3–4 m² m⁻² (Anderson et al., 2004; Birky, 2001; Fassnacht et al., 1997; Wang et al., 2005). Furthermore, wood biomass generally becomes decoupled from LAI after a given stand age, since wood biomass continues to increase after canopy closure. Therefore, a direct relationship between spectral reflectance and forest biomass can only be observed in cases where LAI is both low and correlated to biomass. However, in most forests, LAI reaches values higher than 4 m²/m², and many factors other than stand age can degrade the correlation between LAI and forest biomass (e.g. seasonal variations of LAI, effects of tree density and canopy cover, variation of LAI caused by pest infections or silvicultural practices, etc...). This has limited the use of optical imagery to structurally simple forests, with low biomass, like coniferous forests (e.g. Ardo, 1992; Lu, 2006; Trotter et al., 1997). These methods rely mostly on local calibration and their usefulness in large areas or other contexts is difficult to assess (Curran, 1980; Labrecque et al., 2006).

Despite these limitations, optical images have proved to be interesting if they were combined with other data, like bioclimatic data (Baccini et al., 2004), stand age (Gebreslasie et al., 2010; Zheng et al., 2004), lidar or radar data (Latifi et al., 2010; Lefsky et al., 2005; Walker et al., 2007), or forest inventories data (Wulder et al., 2008). High spatial resolution images allow the use of textural (Proisy et al., 2007; Sarker and Nichol, 2011) or other ecosystem specific methodologies (Leboeuf et al., 2007) for biomass retrieval. Coarse spatial resolution optical satellite images like AVHRR or MODIS, or medium spatial resolution like Landsat or SPOT, have been acquired routinely for years, and their acquisition will continue in the future. Multi-date analyses of surface reflectance allow the examination of changes in forest surface (Chirici et al., 2011; Goward et al., 2008). In addition, the temporal variation of the radiometry can contain new information about forest characteristics, and in particular aboveground biomass (Dong et al., 2003; Powell et al., 2010). Indeed forest growth, and therefore biomass accumulation, is a cumulative function of the net production over time, which is linked to the total amount of absorbed photosynthetically active radiation from the planting date, and therefore could be linked to vegetation indices like NDVI cumulated over time. This has been successfully used since the 80's for herbaceous or agricultural ecosystems (Goward et al., 1985; Nouvellon et al., 2000; Tucker et al., 1981, 1985).

In this study, we analyzed the potential of using *continuous reflectance time-series* for the estimation of forest biomass. We used the 250 m resolution MODIS data, and therefore an un-mixing procedure was necessary to estimate stand-scale reflectances, that were further interpolated to produce continuous time-series. We focused on fast-growing *Eucalyptus* forest plantations in Brazil, which have many advantages for this type of analysis: forest inventories are precise and numerous, stands are large and homogeneous, and short rotations (6 years) enable the monitoring of reflectance since the planting date. In addition, reliable biomass monitoring is of prime significance in *Eucalyptus* plantations as (i) they are now a major worldwide tree crop covering more than 20 million ha globally (of which 4.2 million ha in Brazil, Iglesias-Trabado et al., 2009), and produce large amounts of the wood used by industry for cellulose and charcoal production; (ii) they are highly dynamic both in terms of growth (e.g. 40 m³/ha/yr on average in Brazil, ABRAF, 2009) and expansion (e.g. 300000 hectares of new plantations established per year in Brazil, Iglesias-Trabado et al., 2009); and (iii) they are candidates for mitigation scenarios (Cerri et al., 2010) such as Clean Development Mechanisms.

Our aim was to provide a cost-effective and easily applicable solution for the estimation of *Eucalyptus* biomass, both at the individual stand level and at larger scales. We sought to attain sufficient precision for the analysis of biomass increments at a yearly time step. In this study, we approached biomass through its proxies, volume and dominant height,

which are classically derived from forest inventories. We developed a Direct Remote Sensing approach (Goetz et al., 2009) uniquely based on the use of continuous time series of moderate resolution satellite data, and compared the results with those obtained by including additional, easily accessible bioclimatic variables. Our procedure can be summarized in three steps: 1) unmixing of the MODIS pixel reflectance to get stand-scale NDVI (Normalized Difference Vegetation Index) time-series; 2) estimation of stand age through the analysis of the NDVI time-series; and 3) application of linear and nonlinear regressions to variables derived from the NDVI time-series and bioclimatic variables, and comparison of estimated volumes and dominant heights of individual forest plots with values extracted from a large dataset of field inventories.

2. Material and methods

2.1. Study area

2.1.1. Plantation description

We studied the *Eucalyptus* plantations currently managed for pulpwood production by the International Paper do Brasil company in São Paulo state, south-eastern Brazil (Fig. 1). These fast-growing stands are mainly planted with clones or seedlings of *E. grandis* (W. Hill ex Maiden) x *E. urophylla* (S.T. Blake) hybrids and are currently harvested every six to seven years. In typical plantation management, soil is prepared with mechanical sub-soiling a few months after each clearcut harvest. New cuttings or seedlings are planted in rows at a density of ~1300 trees ha⁻¹ and fertilized. Chemical weeding is carried out during the first year of growth, resulting in the absence of weeds or understory in these plantations. Few stands (less than 1% of the surface) are managed as coppice. The same management practices (e.g. dates of harvesting and weeding, genetic material) are usually carried out for the stands in the same local area. The *Eucalyptus* plantations cover 70,000 ha within a 180 × 250 km region which presents contrasted soil types, from relatively unfertile sandy soils to more fertile soils with a higher percentage of clay. Annual precipitation and temperature vary spatially and from year to year, with a regional average of 1400 mm and 20.2 °C, and local spatial variations between 1250–1510 mm and 19.2–21.3 °C. More than 80% of the rainfall occurs during the wet season between October and April. Mean monthly air temperatures range from about 10 °C to 28 °C.

2.1.2. Forest inventory data

The company that manages these industrial plantations carries out three inventories for each rotation, at ages of about 2 and 4 years, and pre-harvest (5–6 years). Inventory plots have an area of 400 m² and are regularly distributed throughout the stand with a density of one plot per 20 ha. An inventory consists in measuring the diameter at breast height (DBH, 1.3 m above the ground) of each tree contained in the inventory plot, the height of a central subsample of 10 trees, and the height of the four trees of largest DBH (dominant trees). The mean height of the four dominant trees defines dominant height (*Hdom*). *Hdom* is used to calculate a site index (which corresponds to estimated *Hdom* at the age of 7 years) on the basis of growth curves established for each genetic material by the company. Site index, basal area and age at inventory date are then fed into a company-calibrated volume projection equation, specific of the genetic material, to estimate stem volume at any given age. The estimated volume is the volume of wood and bark of the merchantable part of the stem that has a diameter of more than 2 cm. The plot-scale volume (*V*) is calculated as the sum of individual tree stem volumes divided by the area of the plot. Plot-scale *Hdom* and *V* are then averaged at stand scale, for each inventory date.

The company has a geographic information system with the delineation of all the stands. Stand attributes are recorded in the database, like the planting date, soil type, stocking density, etc. All small trials, and stands with other species than *Eucalyptus*, were discarded.

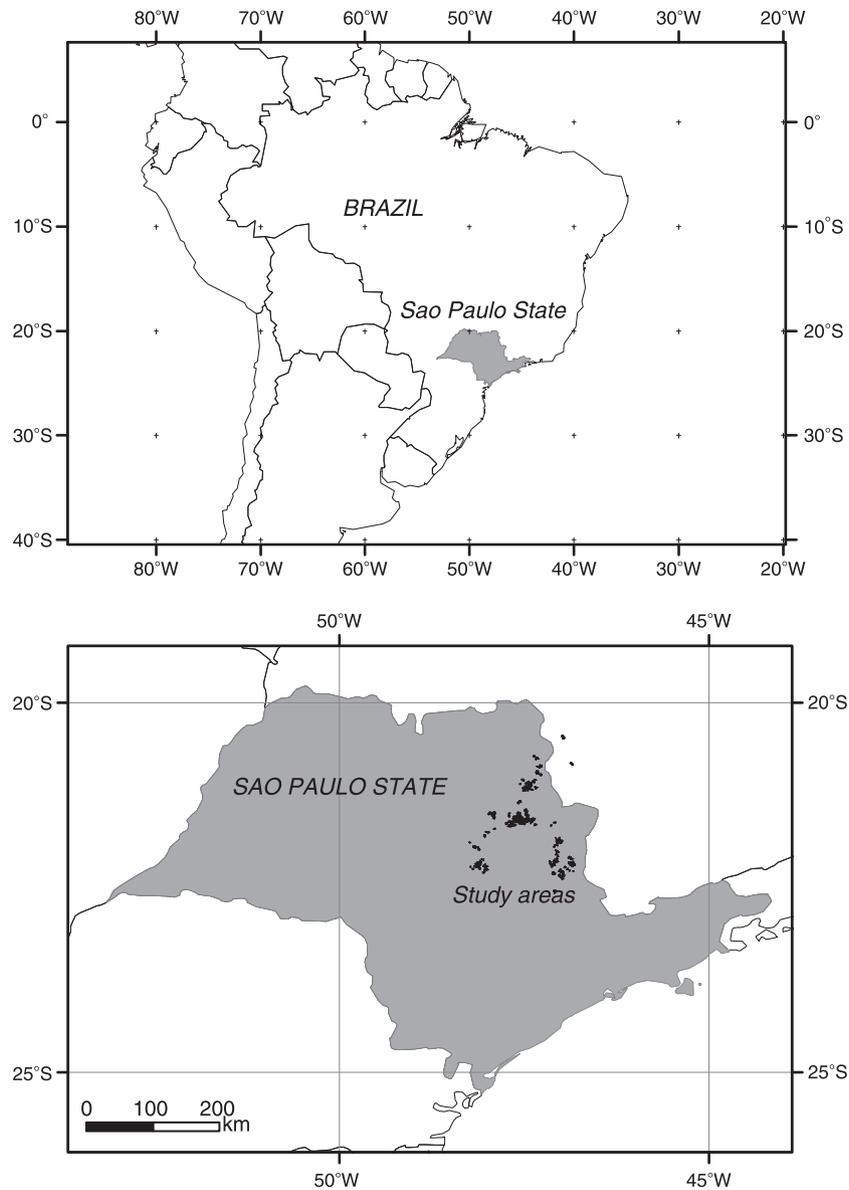


Fig. 1. Location of *Eucalyptus* plantations used in this study, in Sao Paulo State, Brazil.

The average size of remaining stands was 30.1 ha and their median size was 25.4 ha. The largest stands had a surface of about 130 ha. The soil types recorded in the database by the company were grouped into 9 main classes : AQ1 and AQ2 (Arenosols (FAO, 1998), >92% and 85–92% sand in the surface horizon, respectively); LE1, LE2 and LE3 (Rhodic Ferralsols, >60%, 40–60% and <40% clay in the surface horizon, respectively); LV1, LV2 and LV3 (Xanthic Ferralsols, >30%, 10–30% and <10% clay in the surface horizon, respectively); PV (Haplic Acrisols).

2.1.3. Bioclimatic data

WorldClim data (Hijmans et al., 2005) were used to spatially assess main bioclimatic variables. WorldClim database provides monthly precipitations, and mean, minimum and maximum temperature maps with a resolution of 30 arc sec (~1 km resolution) for the period 1950–2000. These maps were generated with data from a network of more than 4000 weather stations worldwide (see Fig. 1 of Hijmans et al., 2005), and interpolated spatially using elevation from SRTM and latitude as covariables. We used the bioclimatic dataset (version 1.4), which are maps of climatic variables calculated from monthly precipitations and

temperature maps. The bioclimatic variables calculated in WorldClim capture annual and intra-annual variations of temperature and precipitation (<http://www.worldclim.org/bioclim>), and are generally used to link climate with vegetation classification or biogeochemical functioning (Sesnie et al., 2008). For each stand of the GIS database, we used the bioclimatic value from the 1 km² pixel that contained the center of the stand. All 19 bioclimatic variables given in WorldClim, which encompass a large range of climatic indicators, were included in our dataset to avoid a subjective *a priori* selection of the variable (Table 1). The regression algorithms we used are able to deal with large numbers of variables.

2.2. Remote sensing data and methods

2.2.1. Satellite images

We used the MODIS/Terra MOD13Q1 products (Vegetation Indices 16-Day L3 Global 250 m, Collection 5), which contain, among other bands, 16-day red and near-infrared reflectances corrected for atmospheric effects. Sun and viewing angles are also given, as well

Table 1

List of all variables calculated from MODIS NDVI time series, Worldclim and soil maps for use as predictor variables with SR and RF regression models.

Variable	Abbreviation	Mean	Min	Max	std	CV
<i>VARage: Stand age, either from the company's records or estimated from NDVI time series</i>						
Stand age (years)	A1	3.52	1.46	7.77	1.52	0.43
Squared age, Age ²	A2	14.69	2.14	60.37	12.15	0.83
Logarithm of stand age, ln(Age)	A3	1.16	0.38	2.05	0.43	0.37
Squared root of stand age, sqrt(age)	A4	1.83	1.21	2.79	0.40	0.22
Sigmoid function of age, sigm(age)	A5	0.39	0.01	1.00	0.39	1.01
<i>VARNDVI: NDVI variables from stand-scale NDVI time-series</i>						
NDVI at inventory date	N1	0.85	0.66	0.94	0.05	0.06
Cumulative NDVI from planting date to inventory date	N2	974.29	365.38	2130.87	465.86	0.48
Cumulative NDVI from planting date to age one year	N3	202.88	154.30	272.51	22.26	0.11
Cumulative NDVI from planting date to age two years	N4	504.28	421.99	582.83	29.71	0.06
Minimum NDVI of the rotation	N5	0.36	0.18	0.62	0.07	0.20
Maximum NDVI of the rotation	N6	0.91	0.83	0.95	0.02	0.03
Mean NDVI of past wet seasons since planting date	N7	0.84	0.72	0.92	0.04	0.05
Mean NDVI of past dry seasons since planting date	N8	0.72	0.57	0.86	0.05	0.07
NDVI of former wet season	N9	0.88	0.78	0.93	0.03	0.03
NDVI of former dry season	N10	0.77	0.59	0.86	0.04	0.05
<i>VARbioclim: Bioclimatic variables from WorldClim (in °C or mm)</i>						
Annual mean temperature	T1	20.20	19.22	21.36	0.55	0.03
Mean diurnal range (mean of monthly (max temp - min temp))	T2	11.63	11.00	12.31	0.44	0.04
Isothermality (ratio between mean monthly temperature and annual temperature range)	T3	0.64	0.63	0.68	0.02	0.02
Temperature seasonality (standard deviation)	T4	21.47	17.58	23.18	1.42	0.07
Max temperature of warmest month	T5	27.91	27.02	29.30	0.56	0.02
Min temperature of coldest month	T6	9.96	8.96	10.95	0.47	0.05
Temperature annual range (T5-T6)	T7	17.95	16.80	18.90	0.60	0.03
Mean temperature of wettest quarter	T8	22.41	21.53	23.41	0.44	0.02
Mean temperature of driest quarter	T9	17.21	16.22	18.53	0.68	0.04
Mean temperature of warmest quarter	T10	22.43	21.62	23.50	0.44	0.02
Mean temperature of coldest quarter	T11	17.13	16.21	18.30	0.59	0.03
Annual precipitation	P1	1377.71	1255.00	1511.62	65.35	0.05
Precipitation of wettest month	P2	253.49	230.00	274.62	11.44	0.05
Precipitation of driest month	P3	22.15	17.00	25.00	2.64	0.12
Precipitation seasonality (coefficient of variation)	P4	71.01	69.00	74.00	1.45	0.02
Precipitation of wettest quarter	P5	692.80	634.00	761.62	32.51	0.05
Precipitation of driest quarter	P6	81.63	63.25	92.00	9.37	0.11
Precipitation of warmest quarter	P7	639.26	576.00	735.09	35.20	0.06
Precipitation of coldest quarter	P8	97.53	73.81	114.00	10.50	0.11
<i>VARsoil: soil type</i>						
Soil type, classified as the main soil types: AQ1,AQ2, LE1, LE2, LE3, LV1, LV2, LV3, PV	S1					

as a Pixel Reliability criterion which classifies pixels according to their quality. All data from the beginning of year 2000 to end of 2009 were downloaded.

In addition to the MODIS data, we used high resolution images to test the validity of our MODIS pixel unmixing procedure (described below). We obtained 31 cloud-free China-Brazil Earth Resources Satellite CCD sensor (CBERS-2 and CBERS-2B CCD) images, for dates between 2003 and 2008 and for two sub-regions of the study area (representing about 7700 ha of plantations). CBERS-2 digital counts were converted to MODIS-like reflectance as described in [le Maire et al. \(2011\)](#). 20 m resolution CBERS reflectances were then averaged per stand using their polygon delineation. These stand red and NIR reflectances based on high resolution images were used as a “best reflectance estimate” which we compared to stand reflectances derived from MODIS unmixing.

2.2.2. Unmixing, filtering and interpolating NDVI time-series

MODIS red and NIR reflectances had a 250 m resolution, which rose the question of the use of such data at the scale of individual stands, whose sizes vary from <5 to several tens of hectares ([le Maire et al., 2011](#); [Marsden et al., 2010](#)). A possibility was to un-mix MODIS reflectance, in order to determine the contribution of different stands to the reflectance of each pixel.

Different methods exist for the downscaling of coarse-resolution images, ranging from very simple techniques to more complex data-

fusion approaches ([Zurita-Milla et al., 2009](#)). We implemented a simple but yet efficient linear mixing model. This model assumes that the reflectance of a large pixel is the result of the linear mixing of the reflectance of its components, weighted by their fractional coverage of the pixel. In our case, the reflectance of a MODIS pixel is the result of the linear mixing of the reflectance of the stands that partly cover the pixel:

$$\rho'_{i,\lambda} = \sum_{j=1}^n f_{ij} \times \rho_{j,\lambda} + \varepsilon_{i,\lambda} \quad (1)$$

$$\sum_{j=1}^n f_{ij} = 1 \quad 0 \leq f_{ij} \leq 1 \quad (2)$$

with $\rho'_{i,\lambda}$ the reflectance of the i^{th} MODIS pixel in the spectral band λ , $\rho_{j,\lambda}$ the reflectance of the j^{th} stand in the spectral band λ , f_{ij} the fractional area of pixel i that is covered by stand j , n the total number of stands, and $\varepsilon_{i,\lambda}$ the residual error.

The main hypotheses underlying Eq. (1) are: i) the reflectance of a stand in a given spectral band is homogeneous in space, ii) the positioning of stand polygons relative to MODIS pixels, considered as squares, is perfect and independent of the spectral band, and iii) the stand cover is spatially continuous, i.e. the reflectance of any small road between stands is neglected.

Considering all stands and pixels, the linear equation (Eq. 1) was converted into a system of N equations (with N the number of MODIS pixels considered), written in matrix-vector notation:

$$R'_\lambda = F \times R_\lambda + \varepsilon_\lambda \quad (3)$$

with R' the vector of the reflectance of all MODIS pixels (size N), R the vector of the reflectance of all stands (size n), F the matrix of fractional cross areas between pixels and stands, of size (N,n) , and ε the vector of residual error (size N). R' , R and ε were a function of the spectral band λ .

Generally, linear unmixing is applied to find the fractional coverage matrix F , knowing R' and with prior knowledge of R (called the end-members vector). In our case, the aim was to determine R based on R' and prior information on F . There was no exact value of R , but a best estimate of R can be found which gives the smallest squared distance between R' and $F \times R$, i.e. the smallest squared error ε . For this purpose, we used the property of the Moore-Penrose pseudoinverse of F , called F^+ (Penrose, 1955). The product of F^+ and R' is the unique solution which minimizes the error. The result of this procedure highly depends on the conditioning of the F matrix. If the matrix is ill-conditioned, the results are very poor. In the case of our *Eucalyptus* plantations, the stand sizes are large compared to a MODIS pixel size, therefore i) the system is overdetermined (many more pixels than stands), and ii) the F matrix is generally well-conditioned, provided small stands and border pixels are filtered out. By trial and error, we found that this filtering had three different thresholds: i) stand size must be larger than 5 ha, ii) a stand must cover at least 40% of at least one pixel, and iii) a pixel must be at least 75% covered with *Eucalyptus* (in one or several different stands). Thresholds ii) and iii) allow a tolerance on pixels at the border of a plantation, which are partially covered with other land-use classes than *Eucalyptus*. In this way, more stands can be kept in the final dataset, but border stands are affected by a larger error. Note however that with this methodology, border errors hardly propagate to other stands. The pseudoinverse method also has technical advantages: F^+ only has to be computed once, since F is considered to be stable with time and spectral bands during the 10 year period, and the calculation of stand-scale reflectance is therefore very fast.

For practical reasons the calculations were separated into 25 zones of grouped *Eucalyptus* plantations. Since a MODIS image is composite, we used the mode of the acquisition date distribution for all MODIS pixels within a zone as a reliable estimate of the date of image acquisition, and the image was kept for applying the unmixing equation only if more than 95% of the pixels within a zone had the best Pixel Reliability value and had a viewing angle lower than 35°.

These simple equations allowed the calculation of the best estimates of stand reflectances, for each MODIS acquisition date. The error of the method was first assessed through the root mean square error of measured vs. simulated MODIS pixel reflectance after stand reflectances R were calculated (RMSEpix, Eq. 4):

$$RMSE_{pix_\lambda} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\varepsilon_{i,\lambda})^2} \quad (4)$$

Finally, the reflectance of the “nearest MODIS pixel”, i.e. the most central pixel of the stand (if the stand contained several pixels) or of the pixel covering the largest fraction of the stand, was extracted for comparison with the stand reflectance obtained using the unmixing methodology. Both methods (linear unmixing and nearest pixel) were compared to the stand reflectance obtained from CBERS-2 images, at the different dates (31 images), with root mean square error calculations (RMSEstand).

Once MODIS NIR and Red reflectances were obtained for each stand and date, the NDVI was calculated as $(NIR-Red)/(NIR+Red)$, and we finally fitted a smoothing cubic spline function to the NDVI time-series of each stand. The parameters of the cubic spline were chosen to get

smoothed values as close as possible to the data, while reducing abrupt discontinuities that were not coherent with vegetation functioning (see Marsden et al., 2010). This smoothing eliminated variations that occurred on a short time scale, not reflecting canopy changes, but rather due to varying viewing angles and residual errors from the atmospheric correction and unmixing procedures. As in Marsden et al. (2010), the smoothing was not very strong since the RMSE of smoothed vs. raw NDVI was lower than 5%. The cubic spline was then used to interpolate the NDVI time series to daily values.

2.2.3. Planting date retrieval

One of the most important variables used in this study was the planting date of each *Eucalyptus* stand. While the planting dates were recorded in the company database, and were therefore easily available, we chose to estimate the planting date from the remotely-sensed NDVI time-series. This was important in view of future applications of our findings, at large scales or in situations where company data is inaccessible or incomplete.

Because MODIS was launched in 2000, it has been possible since ~2007 to use the data to estimate the planting date of most *Eucalyptus* stands currently standing in Brazil, as their rotation length rarely exceeds 7 years.

We used the signal of the clear-cut of the preceding rotation as a robust indicator of planting date. This signal appears unequivocally on the time-series as a sharp decrease of the NDVI, giving a precise estimation of the harvesting date. The planting date of the next rotation can then be inferred reliably by the addition of a certain time-shift (two to three months in these plantations). The planting date itself is difficult to detect directly, because the increase in NDVI immediately after seedlings are planted is generally slow and irregular (short peaks can be due to transitory weed growth for example), so soil or weed NDVI cannot be distinguished from that of young seedlings. In addition, the period of lowest NDVI (after maximum degradation of harvest residues and before planting) is short and can therefore be missed by the 16-day MODIS data, meaning that the minimum NDVI of the time-series does not necessarily reflect the timing of a true minimum just before planting.

The harvesting date automatic retrieval method was based on a moving window of length a , that was split in two. The difference between the average NDVI of the first half and that of the second half was an estimation of the strength of the decrease for the central date of the window. For a given stand j , the window was moved along the NDVI time-series, and the date Dh_j when the difference was largest was recorded as the best estimate of harvesting date.

We calibrated the value of a by testing all values ranging from 3 to 300 days. The date $Dh_j(a)$ was recorded for each value of a and for each stand j . Then we selected the value of a that gave the highest correlation coefficient between Dh and Dp , where Dp were the planting dates recorded by the company. Once a was calibrated, the average difference Lag between Dh and Dp gave the average number of days between the estimated former rotation harvesting date and the present rotation planting date. Planting dates were then estimated as $Dh + Lag$, and tested against company-provided planting dates on a randomly-selected subset of the database (30% of the stands, which were not used to calibrate a).

2.2.4. Variables describing the NDVI time-series

For each stand two smoothed NDVI time-series were prepared, one starting at the estimated planting date and the other starting at the planting date given by the company. This allowed us to test the efficiency and assess the errors generated by using the estimated planting date instead of the real planting date. A number of variables describing the NDVI time-series after planting date were used, and are described in Table 1. These variables were chosen from known properties of the NDVI time-series (Marsden et al., 2010), to describe most of their discriminating features. Two of these variables required that the plantation be more than two years old: the cumulative value of

NDVI from 0 to 2 years (N4), and the maximum NDVI value of the rotation (N6), which generally occurs around two years of age.

2.3. Statistical modeling

Several linear and non-linear regressions were performed on the dataset created (Table 2). The variables we wished to estimate were stand dominant height H_{dom} (m) and stand merchantable wood volume V ($m^3 \cdot ha^{-1}$). We tested various *a priori* variables in the models, from age-only variables to increasingly complex variable combinations (Table 2). All these models were fitted with the same number of observations (i.e. 347 stands with all the information required).

Stepwise multilinear regressions (SR) aimed at automatically choosing a set of predictive explanatory variables among all possible variables, here with a criterion based on the partial F -statistic, with $\alpha = 0.05$ as a threshold for adding or removing a predictor variable. Different models were examined. The first and most simple model SR1 (or SR1') used only estimated (or real) stand age variables as inputs. Since it is well known that the relationship between age and biomass is generally non-linear, we included a set of five different transformations of the age variable: age, age², ln(age), 1/age, and *sigm*(age), where *sigm* is a sigmoid filter function. The second model SR2 (or SR2') consisted in using both estimated (or real) age variables and variables extracted from the NDVI time-series. The last model SR3 (or SR3') further included the WorldClim bioclimatic variables. These stepwise regressions resulted in the selection of a set of explanatory variables with associated linear regression coefficients.

For comparison with the SR models, we performed nonlinear nonparametric regressions using a regression tree algorithm called Random Forest (RF) (Breiman, 2001). Tree-based algorithms recursively split the data into two groups, where the groups are more homogeneous than the unpartitioned data. When the variability within the node is considered sufficient, or when a pre-defined number of classes is reached, the splitting procedure stops. Random Forest combines the tree-based regression and a bootstrap aggregation algorithm (Breiman, 2001). Bootstrap aggregation is a method to average predictions from a collection of bootstrap samples, which provides a valid tool for tuning and improving the accuracy of the predictions (Breiman, 1996). Compared to standard tree-based models, Random Forest is less sensitive to noise in the training data and tends to result in more accurate models. It can deal with large numbers of independent variables, high-order interactions and correlated predictor variables. RF also has the advantage of handling categorical variable as inputs. It was used to estimate the above-ground biomass of forests with high accuracy, e.g. in the United-States (Baccini et al., 2004; Powell et al., 2010), Russia (Houghton et al., 2007) and Africa (Baccini et al., 2008), and it proved to be more efficient in terms of accuracy than other regression approaches (Powell et al., 2010).

Table 2
a priori variables used for the different models. VAR_{age}: age variables (see Sections 2.2.3, 2.3 and Table 1); VAR_{NDVI}: NDVI-based variables (see Section 2.2.4 and Table 1); VAR_{bioclim}: bioclimatic variables (see Section 2.1.3 and Table 1); VAR_{soil}: soil type (Table 1).

Age	Model	Type	Variables
Age from company database	SR1'	Stepwise multilinear regression	VAR _{age}
	SR2'		VAR _{age} , VAR _{NDVI}
	SR3'		VAR _{age} , VAR _{NDVI} , VAR _{bioclim}
	RF1'	Random forest regression	VAR _{age} , VAR _{NDVI} , VAR _{bioclim}
Age from MODIS estimations	RF2'	Random forest regression	VAR _{age} , VAR _{NDVI} , VAR _{bioclim} , VAR _{soil}
	SR1		VAR _{age}
	SR2	Stepwise multilinear regression	VAR _{age} , VAR _{NDVI}
	SR3	Stepwise multilinear regression	VAR _{age} , VAR _{NDVI} , VAR _{bioclim}
	RF1	Random forest regression	VAR _{age} , VAR _{NDVI} , VAR _{bioclim}
	RF2	Random forest regression	VAR _{age} , VAR _{NDVI} , VAR _{bioclim} , VAR _{soil}

Results from RF1 or RF1' models (Table 2) were compared to results obtained with SR3 or SR3'. The improvement of prediction accuracy achieved by including soil type variables was assessed by comparing the additional RF2 model to RF1 (or RF2' to RF1').

The RF algorithm acts as a black box. However several indices were computed to allow the exploration of the model itself: the mean of squared residuals and the percentage of variance explained for the bootstrap sample predictions (named "out-of-bag", Breiman, 2001), and the importance of each variable which was represented by the "mean increase in accuracy" (%IncMSE). All calculations were carried out with the R package randomForest (Liaw and Wiener, 2002).

All models given in Table 2 were calibrated on 70% of the stands selected randomly, and the remaining 30% were used for testing the predictive capacity of the models. This random selection of 70% of stands was repeated 50 times, and four statistics were calculated each time: the coefficient of determination R^2 and root mean square error RMSE between regression-based and inventory-based estimates of stand wood volume and dominant height. These 50 pairs of R^2 and RMSE estimates were presented as box-plots, and their median values served as overall estimates of the model performance. The final equations for SR models, given in this paper, were calculated with the entire dataset.

3. Results

3.1. Unmixing

The RMSE of measured vs. estimated MODIS pixel reflectance after stand reflectances R were calculated (RMSE_{pix}) were low: 0.0068 and 0.0175 for RMSE_{pix,Red} and RMSE_{pix,NIR}, respectively. The RMSE_{pix} values were stable with time, for the ~130 dates of the MODIS time-series that were retained after filtering (data not shown).

The RMSE_{stand} between stand reflectance obtained from MODIS (unmixing or nearest pixel methods) and stand reflectance from the corrected CBERS-2 high resolution images were lower than 0.008 and 0.017 for RMSE_{stand,Red} and RMSE_{stand,NIR}, respectively (Fig. 2). Stand reflectances were retrieved without bias. These low RMSE_{stand} validated the use of MODIS images to estimate stand-scale reflectance.

The unmixing method led to significantly smaller RMSE and higher r^2 for both bands, compared to the simple "nearest pixel" methodology (Fig. 2). This improvement concerned all stand sizes (not shown). The unmixing methodology was therefore used as a trustworthy and simply applicable method to get *Eucalyptus* stand reflectance from MODIS 250 m resolution reflectance.

3.2. General behavior of *Eucalyptus* NDVI time series

Eucalyptus NDVI time series presented a typical pattern of variations, which can be explained by the management and phenology of the plantations. Fig. 3 shows the NDVI time-series of a single stand, together with the cubic spline smoothed curve. The harvesting date of the previous rotation is clearly visible, with a sharp drop of NDVI values. Fig. 3 presents the averaged and standard deviation of NDVI-smoothed time-series of all stands as a function of time, starting on the 1st January of the year of planting. The NDVI starts at values close to soil values, then rises very rapidly during the two first years, up to a maximum value around the second year. The inter-stand variability during this period is very high because of planting dates that occur progressively throughout the first year. After the second year, the NDVI remains high, with a noticeable seasonal decline during dry seasons due to leaf shedding and slowed leaf production (le Maire et al., 2011). The maximum and minimum annual NDVI reduce from the second year to the end of the rotation, which is represented by dashed lines in Fig. 4. All these characteristics resulted in the definition of a series of variables describing the NDVI time-series of a stand. These variables are listed in Table 1,

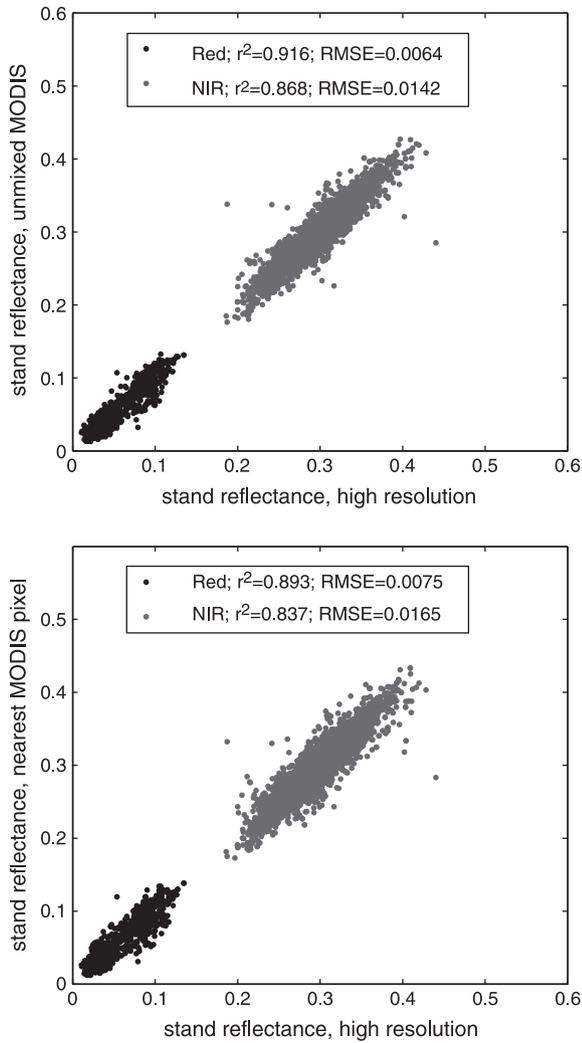


Fig. 2. MODIS estimation of stand reflectance compared to high resolution CBERS stand reflectance, for Red (black dots) and NIR (gray dots) reflectance. MODIS stand reflectance estimations come either from the unmixing methodology (top) or from the nearest pixel methodology (bottom).

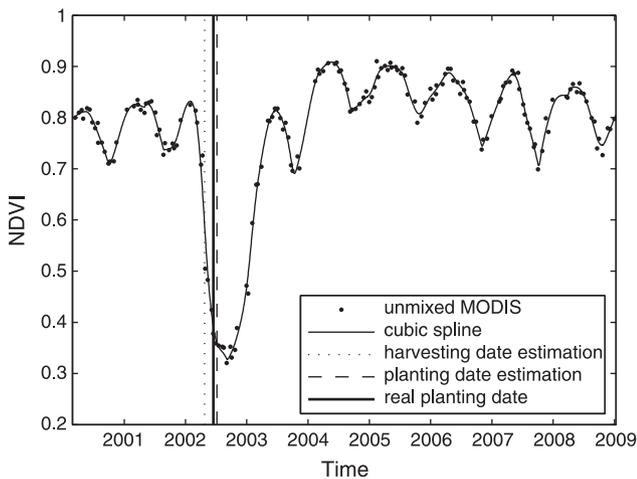


Fig. 3. Example of a stand-scale MODIS NDVI time series obtained after the unmixing procedure and smoothed by a cubic spline function. The harvesting date of the former rotation and the planting date estimated as described in the text are shown together with the real planting date from company records.

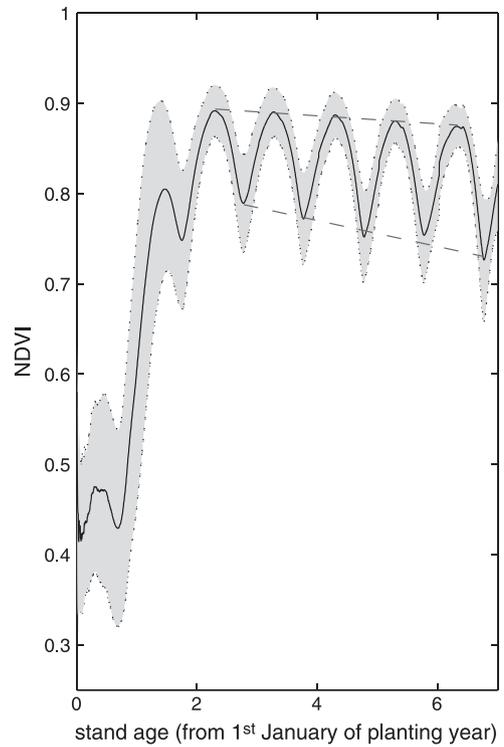


Fig. 4. Typical NDVI age course of a *Eucalyptus* stand. The continuous line is the averaged NDVI-smoothed time-series of all stands as a function of time, and the dotted line delimiting the gray area is \pm their standard deviation. Here the time axis starts on 1st January of the planting year. Dashed lines show the linear fit of annual minimum and maximum NDVI values.

together with their range of variation and average values across all stands.

3.3. Planting date estimation

The algorithm allowed the estimation of the planting date with a root mean square error of less than 40 days. Fig. 5 shows the comparison of the estimated harvesting date D_h with the planting

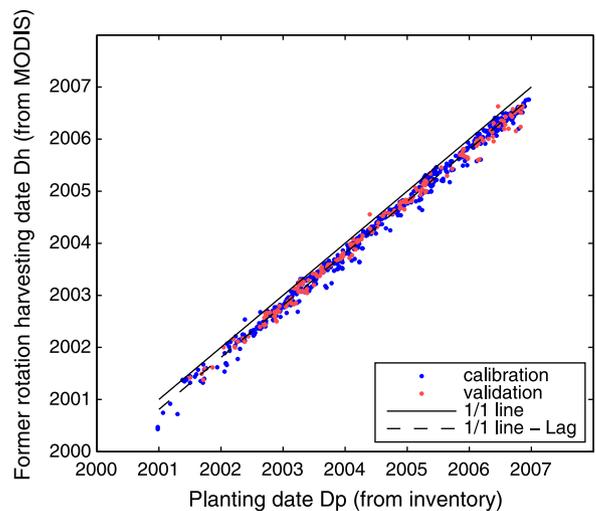


Fig. 5. Former rotation harvesting date retrieved with MODIS NDVI time series compared to planting date from company records. The lag was estimated to be 73 days, and the RMSE once corrected for this lag was 35.5 days.

date D_p , with the best value of the window size $a = 194$ days. Values of a comprised between 160 and 240 gave similar results. The correlation was very high both for the calibration and test dataset ($r^2 > 0.99$). The average difference Lag between D_p and D_h was 73 days, which can be interpreted as the average number of days between the harvesting date of the former rotation and the planting date of the current rotation. After the planting dates were estimated by adding Lag to D_h , the RMSE of the estimation was 35.5 days on the calibration dataset and 37.9 days on the test dataset. A visual example of the estimation of the planting date is given in Fig. 3 for one particular stand.

Fig. 6 shows the box plot of the delay ($D_h - D_p$) for each month, whatever the year. The residuals are fairly well distributed around the mean: the delay between harvesting and the following planting date did not vary seasonally.

3.4. Estimations of volume and dominant height

The coefficient of determination and RMSE of the prediction of stand volume and dominant height when using the planting date from inventory or when using the planting date estimated from the NDVI time-series were not significantly different. The results for volume estimation are presented in Fig. 7, and a very similar pattern of model performance held for H_{dom} (not shown). For instance, the estimation of wood volume (V) gave a RMSE of 24.5 and 26.3 $m^3 ha^{-1}$ with SR3' (using the company planting dates) and SR3 (using estimated planting date) models, respectively, and 1.33 and 1.62 m for dominant height (H_{dom}) estimations.

The results of SR and SR' models showed progressive improvement when NDVI and bioclimatic variables were added successively to age variables (Fig. 7). Models SR1 and SR1', which used only age variables as predictive factors, already explained more than 82% and 87% of the variability of V and H_{dom} , respectively. These values increased to 86% and 90% when variables from the NDVI time-series were added, and to 90% and 92% when bioclimatic variables were added. Finally, the RF1 and RF1' models performed similarly to the SR3 and SR3' models. RF1 and RF2 models were not significantly different, i.e. adding the soil type as a new categorical variable did not improve the model results.

Variables selected in the SR models differed slightly (Table 3). The SR1 models included both age and transformed age variables. Although age variables were correlated, this result showed that the age- V and age- H_{dom} relationships were non-linear. For the SR3

models, A (age-related), N (NDVI-related), T (temperature-related) and P (precipitation-related) variables were used.

Fig. 8 presents all the variables classified by order of importance in the RF2 models. The variables that had a high "importance" value (%IncMSE) were similar in the RF2 model for V and for H_{dom} . The cumulative NDVI from planting date to inventory date (N2) was the most important variable for the estimation of V and H_{dom} . This variable encompassed information about age and about the fluctuations of the NDVI time series. All age variables had very similar importance values, since they were only transformed with nonlinear filter functions, and RF is a nonlinear regression method. Similarly to what was observed with SR models, age explained a large part of the V and H_{dom} variability. The cumulative NDVI of the first two years of growth appeared as one of the most important variables. Temperature variables were found to be the next most important bioclimatic variables. These variables referred mainly to the local thermal amplitude and minimal temperatures. Finally, precipitation variables, like the average yearly precipitation, and precipitation of the wettest and driest quarter were also important. Below a threshold of 10% of the importance %IncMSE, we considered that variables had a low explanatory effect. As expected from the comparison of RF2 and RF3 R^2 and RMSE, soil type had a very low importance in both volume and dominant height models.

A predicted vs. measured scatterplot of H_{dom} and V was drawn for illustrative purposes for one of the randomly-selected test sets (Fig. 9). The estimations of H_{dom} were close to the 1/1 line for all values, whereas high volumes were underestimated (above 250 $m^3 ha^{-1}$). The bias was lower with the RF models than with the SR models. RF models showed better results than SR models on the calibration dataset, but similar results on the test dataset. This could mean that the RF model over-fitted the calibration dataset, but without worsening the final results on the test dataset.

4. Discussion

4.1. Unmixing

The MODIS-derived stand reflectance estimations were close to the reference stand reflectances inferred from high resolution images. The remaining discrepancies could be due to certain hypotheses of the methodology that were not completely fulfilled. Firstly, despite company efforts to homogenize tree growth within each plot, the spectral reflectance of a stand may show some heterogeneity at ~100 m scale, if for instance the stand is located on a gradient of soil fertility. The second hypothesis postulates that the positioning of MODIS pixels, considered as squares, and stand polygons is perfect and independent of the spectral band. The georeferencing of stand polygons was good when compared to high resolution images. However, stand polygons were difficult to compare with MODIS images because of their 250 m resolution and MODIS registration errors (Wolfe et al., 2002), and some errors may arise from this. In addition, MODIS pixels are not square and their effective footprint depends on the acquisition geometry (Wolfe, 2006; Wolfe et al., 2002): even if most of the images were taken with viewing angles lower than 20°, some images may have angles between 20 and 35°, and therefore pixels with an effective resolution of more than 250 m. Finally, we assumed that all stands have continuous cover, which is not the case (they can be crossed by roads, small rivers, etc.). These possible departures from our hypotheses highlight the importance of the stand and pixel filtering step of the unmixing procedure. Accordingly, a complete sensitivity analysis of the three filtering thresholds given in Section 2.2.2 should be carried out, to test the validity of the chosen configuration.

Despite the remaining errors, the unmixing procedure was useful as it gave slightly better results than the "nearest MODIS pixel method". Finally, an important and promising feature of the procedure is that, contrary to the user-dependent selection of "nearest MODIS pixels", it is easy to implement objectively, automatically and at large scales.

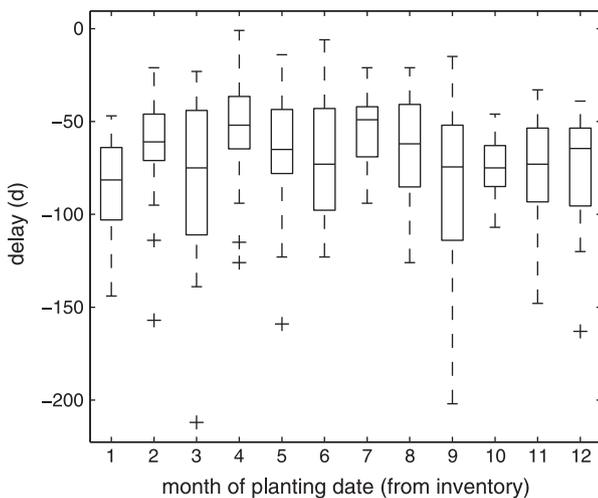


Fig. 6. Boxplots of the delay (days) between the harvesting date, retrieved with MODIS NDVI time series, and the next planting date, given by company inventories, for each month. The boxes of a boxplot represent the median value, upper and lower quartiles, the whiskers are 1.5 times the upper and lower interquartiles range, and crosses are outliers.

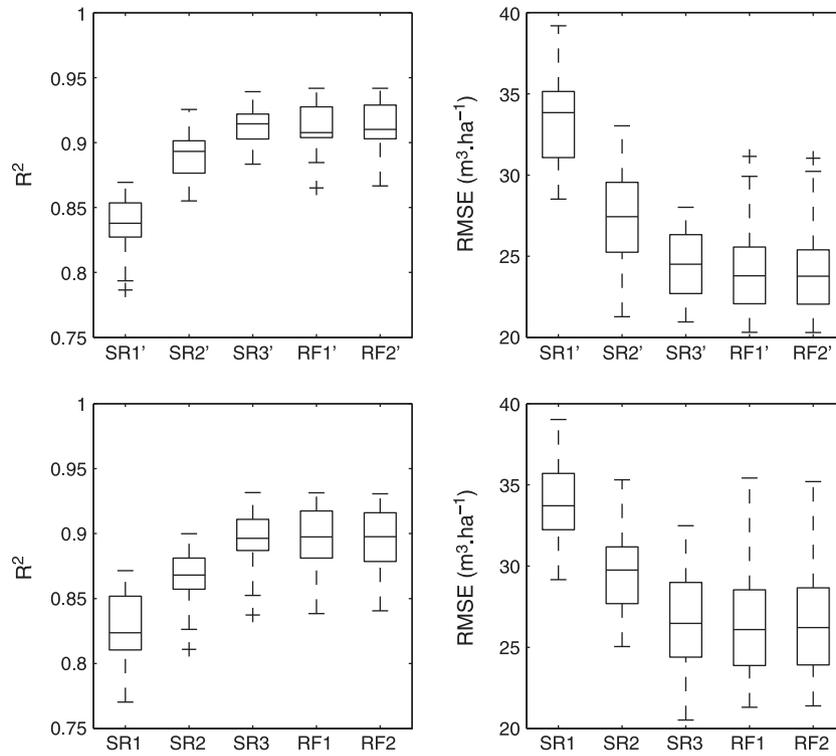


Fig. 7. Volume prediction: boxplots of R^2 and RMSE obtained using the different models described in Table 2. Each model was applied 50 times, each time on a different randomly-selected test sample (30% of the stands).

4.2. Planting date estimation

The planting date retrieval algorithm showed good performance for the purpose of this study, as the use of the MODIS planting date in the models instead of the company planting did not significantly change the results (Fig. 7). However, the RMSE of the prediction of the planting date (35.5 days) is still high and it may require further improvement. There are different possibilities for improving these results: the first one would be to compare the MODIS-derived harvesting date with the harvesting date recorded by the company, to separate the error of harvesting date retrieval from that of the hypothesis of a constant delay between harvesting and planting date (*Lag* parameter). However, the harvesting date was not always recorded in the database, and was subject to some uncertainties (sometimes it was given as a window period for a small region, and moreover the duration of the harvesting operation is

variable between stands). A second possibility would involve the use of daily MODIS data (MOD09GQ product) rather than 16-day composite data. Finally, the combined use of harvesting date estimations with some feature calculated on the moving window related to the initial increase of the NDVI after planting may also improve the planting date estimation.

4.3. Volume and dominant height estimations

The different models presented here proved that it was possible to estimate wood volume and dominant height with a good accuracy, using optical remote-sensing data: the RMSE was around 15% of the mean value for volume and 8% for *Hdom*. Sources of uncertainty in our methodology include the use of continuous NDVI time-series, which are affected by measurement errors due to residual atmospheric effects,

Table 3

Equations of stepwise regression models (SR) for estimating *Eucalyptus* stand volume and stand dominant height. The variables are described in Table 1 and the models are described in Table 2. These equations were obtained on the complete stand dataset.

Model	Equation
SR1'	$V = 97.423 A1 - 6.201 A2 - 123.142$ $Hdom = -82.868 A1 - 290.761 A3 + 652.230 A4 - 10.864 A5 - 541.909$
SR2'	$V = -41.139 A1 + 355.696 A4 + 0.646 N4 - 72.185 N5 - 591.300 N6 + 214.896 N8 - 294.436$ $Hdom = -0.710 A1 + 15.879 A3 + 0.034 N4 - 5.667 N5 - 54.152 N6 + 5.693 N7 + 17.162 N8 + 8.770 N9 + 12.224$
SR3'	$V = 112.016 A1 - 5.705 A2 - 56.634 A5 + 0.470 N3 + 141.490 N8 - 3.899 T2 - 6.773 T6 - 11.135 T9 + 20.031 T11 + 10.440 P4 + 2.137 P6 - 1657.383$ $Hdom = 13.640 A3 + 0.041 N3 - 0.020 N4 + 11.284 N8 - 0.144 T6 - 1.035 T9 + 1.166 T11 - 0.219 P2 + 0.088 P5 - 16.102$
SR1	$V = -3045.625 A1 + 59.496 A2 - 5932.479 A3 + 16402.933 A4 - 13184.306$ $Hdom = -240.376 A1 + 4.645 A2 - 464.587 A3 + 1290.876 A4 - 1028.369$
SR2	$V = 102.385 A1 - 5.036 A2 - 48.513 A5 - 75.514 N1 - 0.569 N3 + 0.944 N4 - 536.308 N6 + 163.648 N8 - 67.619$ $Hdom = -201.411 A1 + 4.377 A2 - 344.875 A3 + 1022.089 A4 - 7.849 N1 - 0.034 N3 + 0.054 N4 - 3.736 N5 - 50.600 N6 + 11.747 N7 + 12.833 N8 + 6.342 N10 - 798.756$
SR3	$V = 86.043 A1 - 4.385 A2 - 0.544 N3 + 0.803 N4 - 578.356 N6 + 106.216 N7 + 163.784 N8 - 0.525 P2 + 0.373 P7 - 0.365 P8 - 157.552$ $Hdom = -116.788 A1 + 2.522 A2 - 184.986 A3 + 570.347 A4 - 6.820 N1 + 0.011 N2 + 0.011 N4 + 8.235 N7 + 7.303 N8 - 14.424 N9 - 0.451 T2 - 0.542 T6 - 0.607 T9 + 0.353 T10 + 0.992 T11 + 0.007 P1 - 495.462$

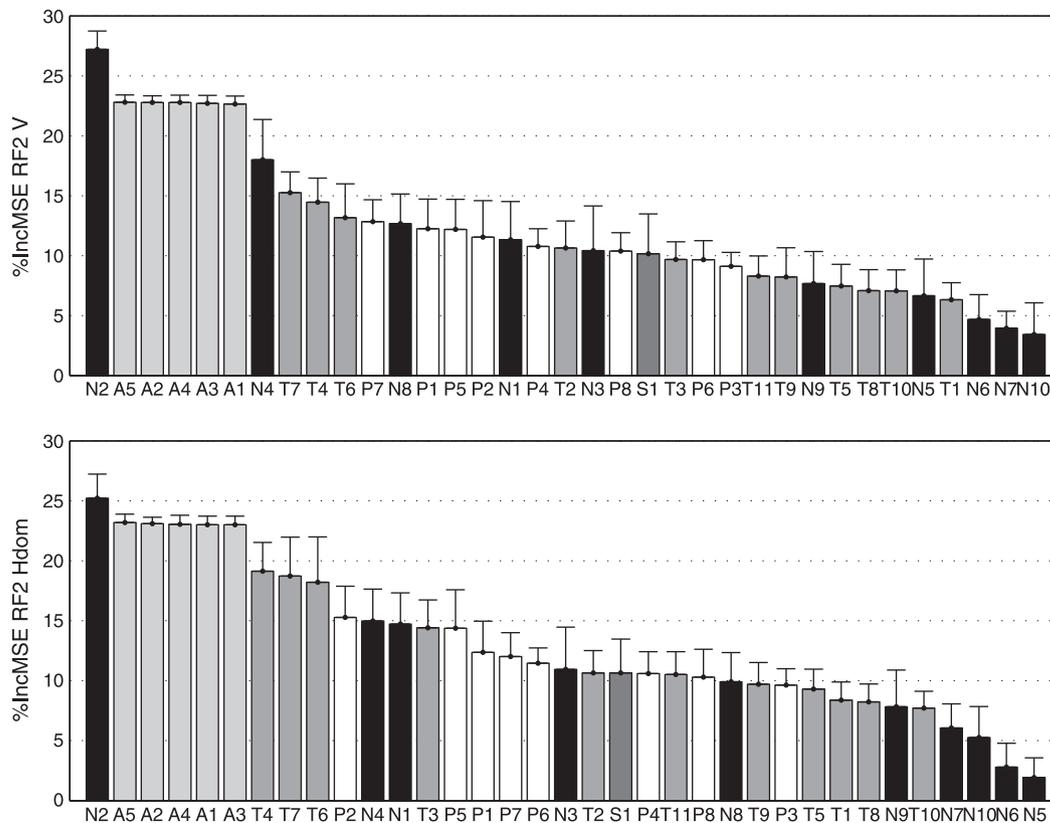


Fig. 8. Classification of the variables by decreasing order of importance in the RF2 models for stand volume (top) and dominant height (bottom). Categories of variables are represented by different levels of gray, and variable descriptions are given in Table 1. The average and standard deviations of the importance (%IncMSE) were calculated on the 50 repetitions of test sample random selection (30% of the stands each time).

viewing geometry variability, and lack of data. Stand reflectances can also be affected by weeds or unusual understory development.

The accuracy of this method, with a RMSE of around 25 m³/ha is comparable to RMSE estimations published for airborne SAR methods that are of around 30–40 m³/ha (Gama et al., 2010). Our method is however largely superior to other passive optical methods. For instance, Figs. 3 and 4 show that NDVI values are not at all related to stand volume or dominant height, that increase continuously with stand age. The accuracy of our method relies mainly on the use of continuous and long NDVI time series. These time series carried very useful information on stand age and on integrated variations of absorbed radiation. In Marsden et al. (2010), it was shown that the NDVI and the fraction of absorbed radiation are closely related on these *Eucalyptus* stands. Therefore, the cumulative NDVI should be roughly correlated with the photosynthetically-active radiation (PAR) absorbed by the stand (e.g. Goetz and Prince, 1996). This explains why the variables representing cumulative NDVI since planting date were among the most important variables in the statistical models. Marsden et al. (2010) also showed, on another smaller dataset, that the cumulative NDVI of the first two years of growth was one of the most important variables for *V* and *Hdom* estimations.

The NDVI alone, even integrated, does not explain all the functioning of the stand. Trees may suffer from periodic water stress or nutrient deficiencies that may not necessarily affect NDVI variations (and therefore the absorbed PAR) but are likely to affect the light use efficiency (Marsden et al., 2010). Management practices (fertilizing, weed control, coppicing) and climate may also vary significantly within a large region like that considered in this study. Some of these differences were taken into account in the models SR3, RF1 and RF2 that included bioclimatic variables. It is difficult to analyze why some bioclimatic variables were selected within the regression models, with an ecophysiological point of view. A process-based modeling approach

would be necessary to disentangle the different effects of climate on plantation growth. However, forest plantation productivity is known to be largely affected by numerous abiotic and biotic variables that can interact. For instance, it is well known that cold temperatures reduce *Eucalyptus* plantation growth (Davidson et al., 2004; Sands and Landsberg, 2002), which probably led to the significance of some variables in the models (like the T6 variable, minimum temperature of the coldest month). Annual precipitation was also variable in the region, and both the rainfall of the wet season and that of the dry season influenced stand volume and dominant height. The dry season influence is quite direct, since water stress directly limits photosynthesis and therefore stand growth (Cabral et al., 2011; White et al., 2009). Wet season rainfall fills the soil water reserve: in these regions with very deep soils, *Eucalyptus* roots can reach depths of more than 15 m at the end of a rotation (Christina et al., 2011). As a consequence, almost all of the incoming rainfall is evapotranspired by the vegetation after canopy closure (at around 2 years of age). It is important to note that the soil type did not bring more information in the RF models. An accurate characterization of soil properties is difficult at the stand level and large differences in stand growth depending on soil types were accurately accounted for by other variables (NDVI-derived, in particular) in RF models. This constraint having been eliminated, the volume and dominant height estimation models can therefore be applied independently of soil types data. The quality of the WorldClim bioclimatic data is dependent on the density of the weather stations used in the interpolations. The region where the study was performed is one of the tropical regions with the highest weather station density for temperature and precipitation data and therefore quality controls were correct (Hijmans et al., 2005).

Our study confirmed that regression trees like the Random Forest algorithm are a flexible and accurate modeling framework for biomass estimation from remote sensing data, as underlined in Powell et al.

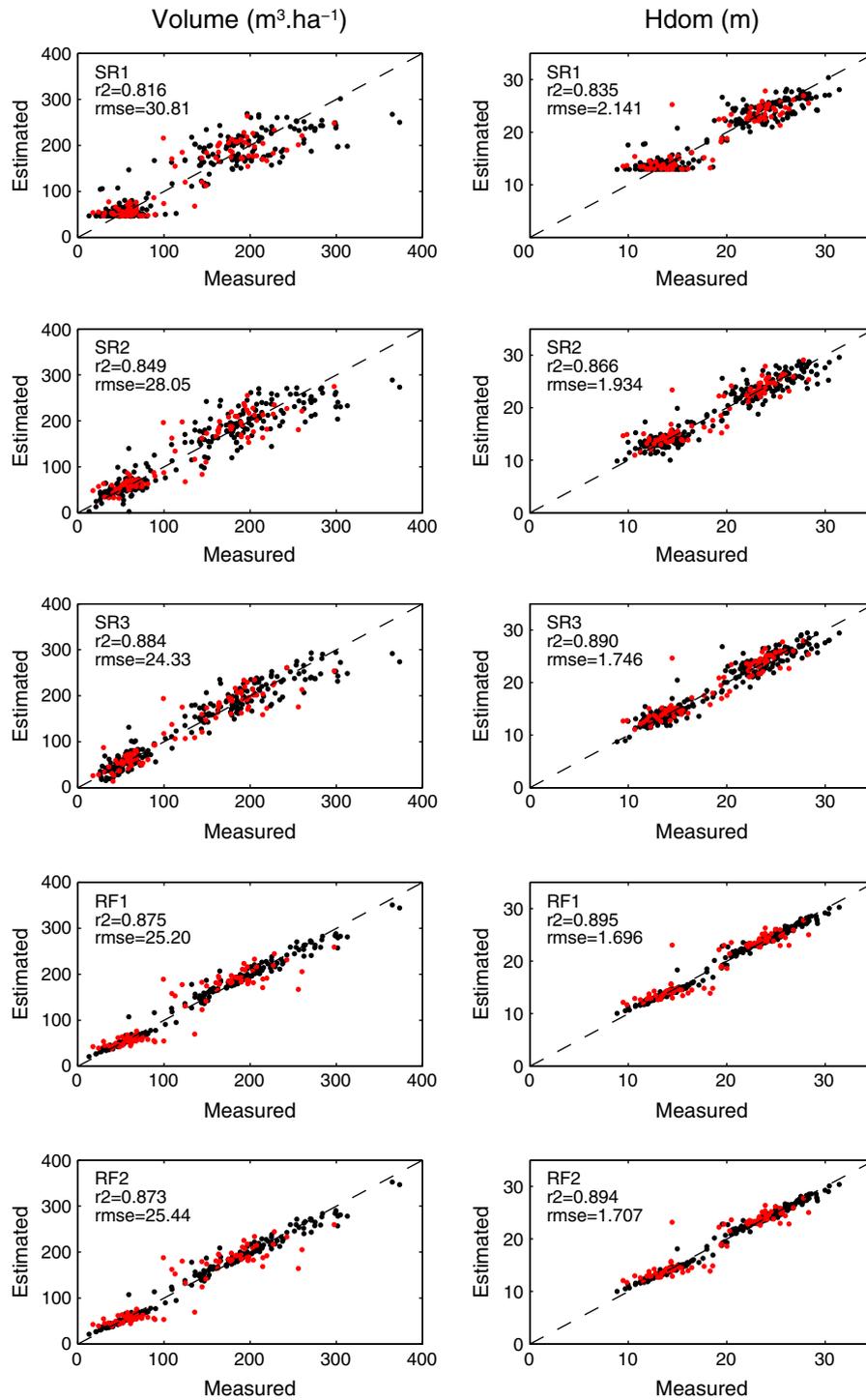


Fig. 9. Illustration of predicted vs. measured volume and dominant height for one of the 50 calibration and test datasets that were used in Figs. 6 and 7. The R^2 and RMSE were calculated for the test sample (red dots).

(2010), Moisen and Frescino (2002) or Blackard et al. (2008). The statistical models we presented could however benefit from further improvements, for example by including additional sources of data in the analysis. The use of local meteorological data from field stations may better explain the interannual variability in growth and in structural characteristics of the plantations than our rough estimates of bioclimatic variables, and for the 2000–2010 period (bioclimatic variables are calculated in WorldClim for the 1950–2000 period). However, model application would be all the more difficult, as meteorological stations are

sparse and their data time-series are often incomplete. An alternative would be the use of other satellite products like TRIMM or MeteoSat data. The NDVI could also be replaced by a more complex soil adjusted vegetation index (SAVI type). It has been shown that the GESAVI, a SAVI-type vegetation index (Gilbert et al., 2002), is more correlated to LAI than NDVI for these eucalypt plantations (le Maire et al., 2011). The GESAVI could be calibrated locally to take into account differences in soil reflectance. Note however that in the NDVI variables, we introduced the minimum NDVI of the time-series (N5), which is a proxy of the soil

NDVI, but this variable was not important in the RF models. Finally, the greatest scope for further improvement of the precision of our methodology could come from the use within a RF framework of complementary sources of temporal satellite data, like radar data, (or other indices using other spectral bands) jointly with MODIS time series, in order to explain the remaining biomass variability that is not captured by reflectance variations.

4.4. Applicability of the method at large scales

As the present study encompassed a wide range of soil types and climatic conditions, our statistical models present good potential for extrapolation purposes, which could be confirmed by tests on other field datasets. Even if the accuracy of these models is probably too low from the perspective of forest managers to replace field inventories, such models can answer large-scale questions about regional *Eucalyptus* wood production and interannual variability. Wood volume can be converted to wood carbon mass provided wood density and carbon content are known. Wood density can be roughly estimated as a function of age (as in de Aguiar Ferreira and Stape, 2009; Marsden et al., 2010), and carbon content can be considered to be approximately constant. Practical estimations of wood biomass are therefore feasible at large and even country-wide scales based on the models developed in this study, at very low cost (all the satellite data used here are freely available) compared to airborne or satellite radar acquisition. An improvement of the unmixing algorithm would be necessary to take into account smaller and border stands that were discarded in the present study. Such improvement could be achieved by further developing the use of multi-date high resolution CBERS images. Their automatic segmentation and classification could result in: i) a grouping of similar stands together based on their homogeneity; ii) the segmentation of other land-use classes needed for unmixing stands at plantation borders; and iii) the detection of possible changes in stand limits or land use. Another approach could be to use CBERS images directly in an unmixing-based data fusion approach (Zurita-Milla et al., 2009).

In addition to its immediate applications for regional carbon accounting or forest management, our large-scale biomass estimation approach can be a useful scientific tool for research work aiming to understand and simulate plantation biogeochemistry. Meteorological and soil conditions interact in a complex way to explain the growth of a forest stand. The use of a stand-scale process-based model that simulates tree growth and water and nutrient cycles at the scale of the stand, taking into account meteorological conditions and soil properties, could be forced with NDVI time-series data to constrain LAI dynamics and therefore improve biomass simulations (see le Maire et al., 2010). Conversely, the use of our large-scale biomass estimates could constrain process-based models in order to better simulate carbon fluxes (Bellassen et al., 2011; le Maire et al., 2005).

The methodology presented here for *Eucalyptus* plantations could be transferred to other types of forests. The main constraints are that the stands have to be large enough, even-aged and homogeneous, and young enough for MODIS NDVI time-series to be available since the planting date (or approximate date of regeneration). The methodology could therefore be applied to other fast-growing forests like poplar, pines, etc.

5. Conclusion

This study proposes a regression-model approach for the large-scale estimation of forest plantation biomass, based uniquely on optical satellite data and other freely accessible information. Contrasting with the generally low efficiency of optical remote sensing for the retrieval of forest biomass, we have shown that the use of continuous NDVI time series starting at planting date gave precise and robust estimates of *Eucalyptus* stand wood volume and dominant height (accuracy of 15 and 8% of their average, respectively). The age of the plantation stands,

which could be obtained from the NDVI time-series themselves, was one of the major variables explaining stand volume and height in these *Eucalyptus* plantations. Additional important variables that significantly improved model efficiency were the cumulative NDVI since planting date or during focused periods, and local bioclimatic characteristics which were obtained from a world database. The resulting models form a cost-effective and accurate method for the mapping of *Eucalyptus* biomass at large scales and for the monitoring of its evolution, which is of high interest for carbon accounting in the forestry sector.

Acknowledgments

The authors acknowledge International Paper do Brasil and in particular Cristiane Camargo Zani, José Mario Ferreira and Sebastião Oliveira, for providing data and technical help. We are grateful to the European Integrated Project Ultra Low CO₂ Steelmaking (ULCOS-contract no. 515960), the Eucflux project (<http://www.ipef.br/eucflux/>), the Usp/Cofecub project (No. 22193PA), and the French Ministry of Foreign Affairs for their financial support.

References

- ABRAF (2009). *Chapter 01, planted forests in Brazil*. Statistical yearbook 2009 2009.
- Anderson, M. C., Neale, C. M. U., Li, F., Norman, J. M., Kustas, W. P., Jayanthi, H., & Chavez, J. (2004). Upscaling ground observations of vegetation water content, canopy height, and leaf area index during SMEX02 using aircraft and Landsat imagery. *Remote Sensing of Environment*, 92, 447–464.
- Ardo, J. (1992). Volume quantification of coniferous forest compartments using spectral radiance recorded by Landsat Thematic Mapper. *International Journal of Remote Sensing*, 13, 1779–1786.
- Baccini, A., Friedl, M. A., Woodcock, C. E., & Warbington, R. (2004). Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters*, 31, L10501.
- Baccini, A., Laporte, N., Goetz, S. J., Sun, M., & Dong, H. (2008). A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environmental Research Letters*, 3, 9pp.
- Baker, D. J., Richards, G., Grainger, A., Gonzalez, P., Brown, S., DeFries, R., Held, A., Kellendorfer, J., Ndunda, P., Ojima, D., Skovseth, P. -E., Souza, C., Jr., & Stolle, F. (2010). Achieving forest carbon information with higher certainty: A five-part plan. *Environmental Science & Policy*, 13, 249–260.
- Bellassen, V., Delbart, N., Le Maire, G., Luyssaert, S., Ciais, P., & Viovy, N. (2011). Potential knowledge gain in large-scale simulations of forest carbon fluxes from remotely sensed biomass and height. *Forest Ecology and Management*, 261, 515–530.
- Birky, A. K. (2001). NDVI and a simple model of deciduous forest seasonal dynamics. *Ecological Modelling*, 143, 43–58.
- Blackard, J. A., Finco, M. V., Helmer, E. H., Holden, G. R., Hoppus, M. L., Jacobs, D. M., Lister, A. J., Moisen, G. G., Nelson, M. D., Riemann, R., Rufenacht, B., Salajano, D., Weyermann, D. L., Winterberger, K. C., Brandeis, T. J., Czaplowski, R. L., McRoberts, R. E., Patterson, P. L., & Tymcio, R. P. (2008). Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment*, 112, 1658–1677.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Cabral, O. M. R., Gash, J. H. C., Rocha, H. R., Marsden, C., Ligo, M. A. V., Freitas, H. C., Tatsch, J. D., & Gomes, E. (2011). Fluxes of CO₂ above a plantation of *Eucalyptus* in southeast Brazil. *Agricultural and Forest Meteorology*, 151, 49–59.
- Cerri, C. C., Bernoux, M., Maia, S. M. F., Cerri, C. E. P., Costa Junior, C., Feigl, B. J., Frazão, L. A., Mello, F. F. d. C., Galdos, M. V., Moreira, C. S., & Carvalho, J. L. N. (2010). Greenhouse gas mitigation options in Brazil for land-use change, livestock and agriculture. *Scientia Agricola*, 67, 102–116.
- Chirici, G., Giuliarelli, D., Biscantini, D., Tonti, D., Mattioli, W., Marchetti, M., & Corona, P. (2011). Large-scale monitoring of coppice forest clearcuts by multitemporal very high resolution satellite imagery. A case study from central Italy. *Remote Sensing of Environment*, 115, 1025–1033.
- Christina, M., Laclau, J. -P., J.L.M.Gonçalves, Jourdan, C., Nouvellon, Y., & Bouillet, J. -P. (2011). Almost symmetrical vertical growth rates above and below ground in one of the world's most productive forests. *Ecosphere*, 2(3) art27.
- Curran, P. J. (1980). Multispectral remote sensing of vegetation amount. *Progress in Physical Geography*, 4, 315–341.
- Davidson, N. J., Battaglia, M., & Close, D. C. (2004). Photosynthetic responses to overnight frost in *Eucalyptus nitens* and *E. globulus*. *Trees - Structure and Function*, 18, 245–252.
- de Aguiar Ferreira, J. M., & Stape, J. L. (2009). Productivity gains by fertilisation in *Eucalyptus urophylla* clonal plantations across gradients in site and stand conditions. *Southern Forests: a Journal of Forest Science*, 71, 253–258.
- Dong, J., Kaufmann, R. K., Myneni, R. B., Tucker, C. J., Kauppi, P. E., Liski, J., Buermann, W., Alexeyev, V., & Hughes, M. K. (2003). Remote sensing estimates of boreal and temperate forest woody biomass: Carbon pools, sources, and sinks. *Remote Sensing of Environment*, 84, 393–410.

- FAO (1998). World reference base for soil resources. *World soil resources report 84*. Rome: FAO/ISRIC/ISSS.
- Fassnacht, K. S., Gower, S. T., MacKenzie, M. D., Nordheim, E. V., & Lillesand, T. M. (1997). Estimating the leaf area index of North Central Wisconsin forests using the landsat thematic mapper. *Remote Sensing of Environment*, 61, 229–245.
- Gama, F. F., dos Santos, J. R., & Mura, J. C. (2010). Eucalyptus biomass and volume estimation using interferometric and polarimetric SAR data. *Remote Sensing*, 2, 939–956.
- Gebreslasie, M. T., Ahmed, F. B., & van Aardt, J. A. N. (2010). Predicting forest structural attributes using ancillary data and ASTER satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 12, S23–S26.
- Gilabert, M. A., González-Piqueras, J., García-Haro, F. J., & Meliá, J. (2002). A generalized soil-adjusted vegetation index. *Remote Sensing of Environment*, 82, 303–310.
- Goetz, S., Baccini, A., Laporte, N., Johns, T., Walker, W., Kellndorfer, J., Houghton, R., & Sun, M. (2009). Mapping and monitoring carbon stocks with satellite observations: A comparison of methods. *Carbon Balance and Management*, 4, 2.
- Goetz, S. J., & Prince, S. D. (1996). Remote sensing of net primary production in boreal forest stands. *Agricultural and Forest Meteorology*, 78, 149–179.
- Goward, S. N., Masek, J. G., Cohen, W., Moisen, G., Collatz, G. J., Healey, S., Houghton, R. A., Huang, C., Kennedy, R., Law, B., Powell, S., Turner, D., & Wulder, M. A. (2008). Forest disturbance and North American carbon flux. *EOS, Transactions, American Geophysical Union*, 89, 105–116.
- Goward, S. N., Tucker, C. J., & Dye, D. G. (1985). North American vegetation patterns observed with the NOAA-7 advanced very high resolution radiometer. *Plant Ecology*, 64, 3–14.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.
- Houghton, R. A., Butman, D., Bunn, A. G., Krankina, O. N., Schlesinger, P., & Stone, T. A. (2007). Mapping Russian forest biomass with data from satellites and forest inventories. *Environmental Research Letters*, 2, 045032.
- Iglesias-Trabado, G., Carballeira-Tenreiro, R., & Folgueira-Lozano, J. (2009). Eucalyptus universalis: Global cultivated Eucalyptus forests Map Version 1.2. Dans *Git Forestry Consulting's; EUCALYPTOLOGICS: Information resources on Eucalyptus cultivation worldwide*. Retrieved from <http://www.git-forestry.com> [October 19th 2009].
- Labrecque, S., Fournier, R. A., Luther, J. E., & Piercey, D. (2006). A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland. *Forest Ecology and Management*, 226, 129–144.
- Latifi, H., Nothdurft, A., & Koch, B. (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: Application of multiple optical/LiDAR-derived predictors. *Forestry*. doi:10.1093/forestry/cpq022.
- le Maire, G., Davi, H., François, C., Soudani, K., Le Dantec, V., & Dufrêne, E. (2005). Modelling annual production and carbon fluxes of a large managed temperate forest using forest inventories, satellite data and field measurements. *Tree Physiology*, 25, 859–872.
- le Maire, G., Marsden, C., Laclau, J. -P., Stape, J. -L., Corbeels, M., & Nouvellon, Y. (2010). Spatial and temporal variability of the carbon budget of tropical eucalyptus plantations assessed using ecosystem modelling and remote-sensing. (<http://www.symposcience.fr/exl-doc/colloque/ART-00002412.pdf>). *International conference on integrative landscape modelling (Landmod2010)*. Montpellier: Edition Quae Symposcience
- le Maire, G., Marsden, C., Verhoef, W., Ponzoni, F. J., Lo Seen, D., Bégué, A., Stape, J. -L., & Nouvellon, Y. (2011). Leaf area index estimation with MODIS reflectance time series and model inversion during full rotations of Eucalyptus plantations. *Remote Sensing of Environment*, 115, 586–599.
- Leboeuf, A., Beaudoin, A., Fournier, R. A., Guindon, L., Luther, J. E., & Lambert, M. C. (2007). A shadow fraction method for mapping biomass of northern boreal black spruce forests using QuickBird imagery. *Remote Sensing of Environment*, 110, 488–500.
- Lefsky, M. A., Turner, D. P., Guzy, M., & Cohen, W. B. (2005). Combining lidar estimates of aboveground biomass and Landsat estimates of stand age for spatially extensive validation of modeled forest productivity. *Remote Sensing of Environment*, 95, 549–558.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2, 18–22.
- Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing*, 27, 1297–1328.
- Marsden, C., le Maire, G., Stape, J. -L., Seen, D. L., Rouspard, O., Cabral, O., Epron, D., Lima, A. M. N., & Nouvellon, Y. (2010). Relating MODIS vegetation index time-series with structure, light absorption and stem production of fast-growing Eucalyptus plantations. *Forest Ecology and Management*, 259, 1741–1753.
- Moisen, G. G., & Frescino, T. S. (2002). Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling*, 157, 209–225.
- Nouvellon, Y., Seen, D. L., Rambal, S., Begue, A., Moran, M. S., Kerr, Y., & Qi, J. G. (2000). Time course of radiation use efficiency in a shortgrass ecosystem: Consequences for remotely sensed estimation of primary production. *Remote Sensing of Environment*, 71, 43–55.
- Patenaude, G., Milne, R., & Dawson, T. P. (2005). Synthesis of remote sensing approaches for forest carbon estimation: Reporting to the Kyoto Protocol. *Environmental Science & Policy*, 8, 161–178.
- Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51, 406–413.
- Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., & Ohmann, J. L. (2010). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, 114, 1053–1068.
- Proisy, C., Couteron, P., & Fromard, F. (2007). Predicting and mapping mangrove biomass from canopy grain analysis using Fourier-based textural ordination of IKONOS images. *Remote Sensing of Environment*, 109, 379–392.
- Sands, P. J., & Landsberg, J. J. (2002). Parameterisation of 3-PG for plantation grown Eucalyptus globulus. *Forest Ecology and Management*, 163, 273–292.
- Sarker, L. R., & Nichol, J. E. (2011). Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sensing of Environment*, 115, 968–977.
- Sesnie, S. E., Gessler, P. E., Finegan, B., & Thessler, S. (2008). Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sensing of Environment*, 112, 2145–2159.
- Trotter, C. M., Dymond, J. R., & Goulding, C. J. (1997). Estimation of timber volume in a coniferous plantation forest using Landsat TM. *International Journal of Remote Sensing*, 18, 2209–2223.
- Tucker, C. J., Holben, B. N., Elgin, J. H., Jr., & McMurtrey Iii, J. E. (1981). Remote sensing of total dry-matter accumulation in winter wheat. *Remote Sensing of Environment*, 11, 171–189.
- Tucker, C. J., Vanpraet, C. L., Sharman, M. J., & Van Ittersum, G. (1985). Satellite remote sensing of total herbaceous biomass production in the senegalese sahel: 1980–1984. *Remote Sensing of Environment*, 17, 233–249.
- Walker, W. S., Kellndorfer, J. M., LaPoint, E., Hoppus, M., & Westfall, J. (2007). An empirical InSAR-optical fusion approach to mapping vegetation canopy height. *Remote Sensing of Environment*, 109, 482–499.
- Wang, Q., Adiku, S., Tenhunen, J., & Granier, A. (2005). On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sensing of Environment*, 94, 244–255.
- White, D. A., Crombie, D. S., Kinal, J., Battaglia, M., McGrath, J. F., Mendham, D. S., & Walker, S. N. (2009). Managing productivity and drought risk in Eucalyptus globulus plantations in south-western Australia. *Forest Ecology and Management*, 259, 33–44.
- Wolfe, R. (2006). MODIS geolocation. *Earth Science Satellite Remote Sensing*, 50–73.
- Wolfe, R. E., Nishihama, M., Fleig, A. J., Kuyper, J. A., Roy, D. P., Storey, J. C., & Patt, F. S. (2002). Achieving sub-pixel geolocation accuracy in support of MODIS land science. *Remote Sensing of Environment*, 83, 31–49.
- Wulder, M., White, J., Fournier, R., Luther, J., & Magnussen, S. (2008). Spatially explicit large area biomass estimation: Three approaches using forest inventory and remotely sensed imagery in a GIS. *Sensors*, 8, 529–560.
- Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., & Ryu, S. -R. (2004). Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment*, 93, 402–411.
- Zurita-Milla, R., Kaiser, G., Clevers, J. G. P. W., Schneider, W., & Schaepman, M. E. (2009). Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. *Remote Sensing of Environment*, 113, 1874–1885.