STRUCTURAL CHARACTERIZATION OF CANOPIES OF Eucalyptus spp. USING RADIOMETRIC DATA FROM TM/Landsat 5

Ludmila Roque Pacheco1, Flávio Jorge Ponzoni2, Sandra Benfica dos Santos3, Clódis de Oliveira Andrades Filho4, Márcio Pupin Mello5, Rogério Costa Campos6

(received: March 11, 2010; accepted: October 28, 2011)

ABSTRACT: Empirical approaches and, more recently, physical approaches, have grounded the establishment of logical connections between radiometric variables derived from remote data and biophysical variables derived from vegetation cover. This study was aimed at evaluating correlations of dendrometric and density data from canopies of Eucalyptus spp., as collected in Capão Bonito forest unit, with radiometric data from imagery acquired by the TM/Landsat-5 sensor on two orbital passages over the study site (dates close to field data collection). Results indicate that stronger correlations were identified between crown dimensions and canopy height with near-infrared spectral band data ($\rho$ $\lambda$), irrespective of the satellite passage date. Estimates of spatial distribution of dendrometric data and canopy density ($D$) using spectral characterization were consistent with the spatial distribution of tree ages during the study period. Statistical tests were applied to evaluate performance disparities of empirical models depending on which date data were acquired. Results indicated a significant difference between models based on distinct data acquisition dates.

Key words: Spectral characterization, estimate of biophysical data, reflectance of dossiês.

CARACTERIZAÇÃO ESTRUTURAL DE DOSSÉIS DE Eucalyptus spp. MEDIANTE DADOS RADIOMÉTRICOS TM/Landsat 5

RESUMO: Abordagens empíricas e, mais recentemente, abordagens físicas têm fundamentado o estabelecimento de conexões lógicas entre variáveis radiométricas provenientes de dados remotamente coletados e outras de caráter biofísico provenientes da cobertura vegetal. Nesse sentido, foram avaliadas correlações entre dados dendrométricos e de densidade de dossiês de Eucalyptus spp. coletados na Unidade Florestal de Capão Bonito da empresa FIBRIA S/A e dados radiométricos extraídos de imagens do sensor TM/Landsat 5, provenientes de duas passagens (datas próximas à coleta de dados em campo) sobre a área de estudo. Os resultados indicaram que as maiores correlações foram identificadas nas relações estabelecidas entre as dimensões da copa e a altura dos dossiês com dados da banda espectral do infravermelho próximo ($\rho$ $\lambda$), independentemente da data de passagem do satélite. As estimativas da distribuição espacial de dados dendrométricos e de densidade do dossel ($D$) fundamentadas na caracterização espectral foram consistentes em relação à distribuição espacial das idades dos plantios para o período de realização do trabalho. Foram aplicados testes estatísticos para avaliar diferenças de desempenho de modelos empíricos definidos a partir de dados radiométricos de datas diferentes. Os resultados indicaram que houve diferença significativa entre os modelos estabelecidos a partir de dados provenientes de datas diferentes.

Palavras-chave: Caracterização espectral, estimativa de dados biofísicos, canopy reflectance.

1 INTRODUCTION

Establishing empirical relationships between biophysical variables and spectral variables derived from airborne or orbital sensors when studying vegetation has been the subject of investigation ever since remote sensing data first became available (PONZONI; SHIMABUKURO, 2007). This approach gained prominence after discussions about global climate change took place in which CO2 emissions from burning or exploration of forest biomass were the central point of discussion, particularly in tropical regions. Examples include the works of several authors.

---

1Environmental Engineer, MSc in Remote Sensing – Instituto Nacional de Pesquisas Espaciais – Av. dos Astronautas, 1758, Jd. Granja – 12227-010 – São José dos Campos, SP, Brasil – ludmila@dsr.inpe.br
2Forest Engineer, PhD in Forest Sciences – Instituto Nacional de Pesquisas Espaciais – Av. dos Astronautas, 1758, Jd. Granja – 12227-010 – São José dos Campos, SP, Brasil – flavio@dsr.inpe.br
3Agronomic Engineer, MSc in Energy Engineering Sciences – Instituto Nacional de Pesquisas Espaciais – Av. dos Astronautas, 1758, Jd. Granja – 12227-010 – São José dos Campos, SP, Brasil – sandra@dsr.inpe.br
4Geographer, PhD candidate in Geosciences – Instituto Nacional de Pesquisas Espaciais – Av. dos Astronautas, 1758, Jd. Granja – 12227-010 – São José dos Campos, SP, Brasil – clodiss@dsr.inpe.br
5Land Survey Engineer, PhD candidate in Remote Sensing – Instituto Nacional de Pesquisas Espaciais – Av. dos Astronautas, 1758, Jd. Granja – 12227-010 – São José dos Campos, SP, Brasil – mello@dsr.inpe.br
6Agronomic Engineer, PhD in Remote Sensing – Departamento de Ciências Exatas, Escola Superior de Agricultura “Luiz de Queiroz”/ESALQ – Universidade de São Paulo/USP – Av. Pádua Dias, 11 – Cx. P. 9 – 13418-260 – Piracicaba, SP, Brasil – rogerio.c.campos@hotmail.com

Aragão et al. (2005) studied spatial variability of leaf area index (IAF) in the east portion of the Brazilian Amazonia. Foody et al. (2003) evaluated estimate possibilities for forest biomass in two parts of Amazonia by using data from the Thematic Mapper sensor aboard the Landsat 5 satellite (TM/Landsat 5). Steininger (2000) estimated biomass of secondary forest formations in forests of Brazil and Bolivia, establishing empirical relationships between forest biomass and radiometric data acquired by the CHRIS-PROBA sensor in Tapajós national forest.

The fragility of empirical relationships in studies of this kind is well acknowledged. Schlerf and Atzberger (2006) argued that, where quantitative approaches are used involving use of remote sensing data in vegetation studies, the application of empirical or physical models can be explored. Empirical models usually relate Bidirectional Reflectance Factors (FRB) or Vegetation Indices (IVs) to one or more biophysical variables by defining statistical regression models. However, statistical relationships derived from such models are specific to each sensor, study region and sampling condition specified, in addition to varying in time and space (COLOMBO et al., 2003), all of which validate the application of such models only under the local experimental or environmental conditions they were defined for (ASNER et al., 2003).

A physical approach is based on the application of models grounded on radiative transfer theory, assumed to describe spectral variations of plant canopies as a function of their structural characteristics and spectral properties of the leaves and the soil (GOEL, 1988). According to Houborg et al. (2007), radiative transfer models are useful for explaining the transfer and interactions of electromagnetic radiation in the interior of canopies. Grounded on physical laws, it is possible to establish correlations of biochemical and/or biophysical variables with canopy reflectance. These correlations are said to be ‘direct’ when spectral properties (for instance, FRB) are defined as a function of biochemical and/or biophysical parameters, and said to be ‘inverse’ when biophysical and/or biochemical parameters are defined as a function of observed spectral values (SCHLERF; ATZBERGER, 2006).

Despite there being arguments in favor of physical models, whether direct or inverse, the application of these models is not trivial and deserves different types of adaptation and consideration depending on the canopy being studied (CAMPOS, 2008). Therefore, empirical models are still being used, and their results have varying degrees of success and limitations on the scope of their application.

Canavesi and Ponzoni (2007) used empirical modeling to estimate spatial distribution of wood volume in stands of *Eucalyptus* spp. The authors evaluated the effect of topography on the values of bidirectional reflectance factors extracted from TM/Landsat-5 imagery and concluded that, even empirically, it was possible to estimate wood volume in the eucalyptus stands based on spectral properties.

One of the most desirable applications of remote sensing techniques as far as vegetation studies are concerned is estimating dendrometric and structural data of forest canopies using radiometric data. Specialists in this field, however, are for the most part unfamiliar with the potential and the limitations of quantitative approaches via remote sensing techniques, there being considerable uncertainty or even false expectations as to what is indeed possible and feasible.

The objective of this study is to evaluate the possibility of estimating dendrometric and canopy density parameters using empirical models that correlate them with radiometric data from TM/Landsat-5 imagery.

### 2 MATERIAL AND METHODS

#### 2.1 Study site and field data collection

The study site (located between latitudes 23°46’30” and 24°1’26” south and longitudes 48°15’4” and 48°30’2” west) consisted of eucalyptus stands owned by FIBRIA S/A, located in the municipality of Capão Bonito (Figure 1), São Paulo state. According to geomorphologic mapping of São Paulo state performed by the Instituto de Pesquisas Tecnológicas do Estado de São Paulo (1981), Capão Bonito sits in a region known as São Paulo Peripheral Depression, in Paranaíba Zone. The local relief consists of hills with convex and tabular tops, local altitude ranges between 600m and 700m and local declivity ranges between 10% and 20%.

Field data were collected between May 4 and May 8, 2009. During that period, 30 plots were visited at differing ages (1 to 9 years) and with differing spacing between trees (3 x 1.5m, 3 x 2m and 3 x 3m) in order to obtain canopy data with dissimilar structural characteristics. It should be noted that all plots differed as to clone origin so as to ensure specificity of biophysical uniformity at each age.

In each of the 30 plots visited, data were collected from 10 successive trees in the same planting row. Data included diameter at breast height (DAP), total height (H), stem height (Ht), crown radius (Rc) and canopy density (D).
The first tree to be measured in the plot was selected randomly and its exact location was recorded by a GPS (Global Positioning System). The spatial positioning of the first tree served only as a reference for identifying that particular plot in orbital images, since the radiometric data were derived from the arithmetic average of data from a 3 x 3 pixel matrix about such coordinate. 

**DAP** and **Rc** were measured using a diameter tape, while **H** and **Ht** were measured using a Bitterlich hypsometer with a 20m-distance scale. Crown height (**Hc**) values were estimated using the differences between **H** and **Ht** values. All the above parameters were represented by average values computing the 10 trees. **D** values were estimated using a densitometer, as described by Souza (2009), and served to estimate the percentage of canopy cover in the plots. In each patch, the sampling procedure to estimate **D** consisted of doing a triangle-shaped circuit. Readings were taken every 3 meters, with the densitometer facing upward and reporting whether the sampling point intercepted the canopy or not. Around 100 meters were done on each side of the triangle circuit, to a total of 33 sampling points on two sides, and 34 points on the third side. Figure 2 illustrates the triangle-shaped circuits within each plot. The first densitometer reading point coincided with the site where the first tree was randomly selected for the sampling of biophysical parameters, in such way that the last reading would coincide with the starting point. The final estimate of canopy density corresponds to the number of affirmative points divided by the total number of points being sampled. For instance, assuming that in one patch 75 points correspond to presence and 25 points correspond to absence of canopy elements, then the percentage of canopy density is 75%.

**2.2 Orbital imagery**

Images from path/row 220/77 of the TM/Landsat-5 sensor were selected, in six spectral bands covering the visible (bands 1, 2 and 3), near-infrared (band 4) and mid-infrared (bands 5 and 7) regions on two different dates, as close as possible to the dates of field data collection, namely 22/Apr/2009 and 24/May/2009.
The reason for selecting two satellite passage dates was to check the degree of consistency between the established correlations independently, in other words, correlations were established between field data and radiometric data from satellite images on two occasions and the differences between the established correlations were then assessed. Additionally, a statistical procedure was adopted to assess the differences between the defined regression models relating dendrometric data and spectral data from images acquired on the two dates.

The digital numbers (NDs) of the images acquired on the two dates were converted to apparent reflectance values by applying equations 1 and 2.

\[
\rho_s = \frac{\pi L_s d^2}{E_{sun}, \cos(\theta_s)}
\]

where:

- \(\rho_s\) = Apparent reflectance;
- \(L_s\) = Radiance at the top of the atmosphere in spectral band \(\lambda\);
- \(d\) = Earth-Sun distance in astronomical units;
- \(E_{sun}\) = Solar irradiance at the top of the atmosphere in spectral band \(\lambda\);
- \(\theta_s\) = Solar elevation angle at the time of image acquisition.

\(L_s\) values were determined by applying equation 2.

\[
L_s = \left( \frac{L_{max} - L_{min}}{QCALMAX} \right) QCAL
\]

where:

- \(L_s\) = Radiance at the top of the atmosphere in spectral band \(\lambda\);
- \(L_{max}\) = Maximum radiance when \(QCAL = QCALMAX\);
- \(L_{min}\) = Minimum radiance when \(QCAL = 0\);
- \(QCAL\) = Values of digital numbers referring to and proportional to \(L_s\);
- \(QCALMAX\) = Maximum digital number when \(L_s = L_{max}\).

\(L_{max}\) and \(L_{min}\), values were derived from Chander et al. (2009) according to each date of image acquisition.

Images converted to apparent reflectance values were then subjected to an atmospheric correction procedure by applying the Second Simulation of the Satellite Signal in the Solar Spectrum (6S) code of radiative transfer (VERMOTE et al., 1997), converting them to surface reflectance values (\(\rho_s\)). Images were then georeferenced based on an image from the ETM+/Landsat sensor, available on the ‘Global Land Cover Facility’ website (http://www.landcover.org/data/). A polynomial of degree 1 was used for georeferencing, and pixel resampling was done by using the nearest neighbor method. Twelve visually distinguishable control points were selected in the two images, a total error of about 0.4 pixels being reached, or 12 m.

2.3 Definition of regression models

Surface reflectance values were extracted from the images by computing the arithmetic average of a 3 x 3 pixel matrix relative to the 10 field-measured trees. This procedure was adopted in all 30 plots visited in the field and each spectral band of each image (on each date), excepting point no. 13 as in the image acquired on 24/May/2009 this point coincided with recently harvested areas, and values were then correlated to dendrometric and \(D\) data using the Pearson correlation coefficient (\(r\)), as suggested by Mello et al. (2008).

Investigations based on the correlation values found led to use of simple linear regression models, computing biophysical values (dendrometric variables and \(D\)) as dependent variables and computing reflectance values associated to the near-infrared spectral band on each of the two acquisition dates as independent variables. For the determination of each model, 2/3 of the field-visited points (20 out of 29) were randomly selected for use in the model fitting procedures. Data from the 9 remaining plots were used later in the assessment of resulting models. Due to limited amount of data, this procedure was repeated 10,000 times, randomly and independently, using software R (R DEVELOPMENT CORE TEAM, 2009) as suggested by Peternelli and Mello (2007). It should be noted that the twenty randomly selected points for fitting each model on one date were the same used for fitting the models on the second date.

Among the 10,000 resulting models, for each dendrometric variable and \(D\) and for each date, the model having the highest \(R^2\) among the 10,000 fitted equations was used for estimating the dendrometric variables and \(D\) and for generating the thematic maps for each date independently.

2.4 Assessment and comparison of regression models

In order to statistically assess possible differences between the predictive capabilities of the models relating to the two image acquisition dates, a proportion hypothesis test was applied using a Gaussian distribution (represented by a standard z normal). An analysis was conducted of the coefficient of determination (\(R^2\)) of all 10,000 fitted models for each dendrometric variable and \(D\) and for each date.
By counting the number of times the $R^2$ obtained through regression fitting for a given variable on the April date was higher than the $R^2$ obtained for the May date, it was possible to conduct the test, based on the hypothesis that, if both models have the same predictive capability, then the proportion of times the $R^2$ relating to the April image was higher than the $R^2$ relating to the May image should be equal to 0.5. The proportion test with the respective hypotheses is given as follows:

1) $H_0: p = 0.5$ (50% of the times the $R^2$ of the fitted models using the April image was better and 50% of the times the $R^2$ of the fitted models using the May image was better, which means that the fitted models for both images have the same predictive capability);

2) $H_1: p > 0.5$ (a proportion higher than 50% indicates that the April image has greater predictive capability).

If a 0.05 significance level is adopted, the $H_0$ hypothesis will be rejected whenever the test statistic (calculated z-value) is greater than the tabulated distribution value at the established significance level (critical z-value = 1.645).

Another statistical procedure was applied to assess the regression models for the two dates as to their consistency in predictive capability. The values referring to the 9 points not used in the model fitting process were then used so as to allow a direct comparison between estimated and observed values. According to Kutner et al. (2005), an excellent way of measuring the actual predictive capability of a regression model is to use it for predicting each new data set and then compute the Mean Square Prediction Error (MSPR). MSPR assessment is done through comparison with the Mean Square Error (MSE). The assessment consists of comparing the MSPR associated to predictions (using the 9 points) and the MSE associated to model fitting (using the 20 points): if the obtained MSPR value is lower than the MSE value, then it is an indication that the predictive capability of the model is unsuitable. If the MSPR value is close to the MSE value, then the model is unbiased and fits well to new data. And lastly, if the MSPR value is higher than the MSE value, it is an indication that the model can make good future predictions when used, for instance, with a new data series (KUTNER et al., 2005).

The comparison of MSPR and MSE was done through the average percentage differences (Equation 3) for the 10,000 pairs of fitted models, for each dendrometric variable and $D$.

$$\text{Difference} = \frac{(\text{MSPR} - \text{MSE})}{\text{MSE}} \quad (3)$$

If the relative difference results in a value close to 0 then it is an indication of similarity and that the model is unbiased, and if it results in a value greater than 1, MSPR is greater than MSE, indicating that the model has good future prediction capability (KUTNER et al., 2005).

3 RESULTS AND DISCUSSION

Correlation coefficients ($r$) were determined between the dendrometric (Figure 3) and structural variables and the $\rho$ values, with results being provided in Table 1.

Table 1 shows that higher $r$ values were found for $\rho_4$ on both dates. The negative $r$ values for the correlations of $\rho_4$ with $DAP$, $H$, and $Ht$ were expected, inasmuch as older stands are expected to provide deeper shade at the top of the canopy and thus decrease the $\rho_4$ values (CANA VESI; PONZONI, 2007). This relationship becomes direct when $Hc$, $Rc$ and Densitometer readings are correlated to $\rho_s$. In other words, the greater the amount of leaves at the top of the canopy, the higher the expected $\rho_s$ values (GOEL, 1988; KUMAR, 1974; PONZONI; SHIMABUKURO, 2007).

Linear regression models were established for the correlations with higher $r$ values, using dendrometric variables and $D$ as dependent variables and radiometric measurements as independent variables. For thematic maps, the model having the highest $R^2$ value was selected, out of the 10,000 models generated for each date and for each dendrometric variable and $D$. Figure 4 provides the thematic maps developed for TM/Landsat-5 images (22/April and 24/May), referring to $DAP$, $H$ and $Ht$.

To facilitate discussions on the thematic maps generated, another thematic map was developed using existing company data referring to spatial distribution of tree ages in the stands, as illustrated in Figure 5.

Through an analysis of the generated thematic maps and helped by the map of spatial distribution of ages across the eucalyptus stands at the time of field data collection, it can be said that, overall, the mapping of dendrometric variables showed consistency. The variation of dendrometric parameters throughout mapping was found to agree with the ages of the stands. In other words, the lowest values of dendrometric parameters $DAP$, $H$, $Ht$ occurred in areas where stands were younger (Figures 4a, 4b, 4c, 4d, 4e and 4f). Consistency was also observed in the other parameters, $Hc$, $Rc$ and $D$, accordingly with lower values the older the stand was (Figures 4g, 4h, 4i, 4j, 4l and 4m), indicating that younger stands have more leaf biomass and thus higher $Hc$, $Rc$ and $D$ values.
Figure 3 – Variation in biophysical parameters collected in the field as a function of age, DAP (a), H (b), Ht (c), Hc (d), Rc (e) and D (f).

Figura 3 – Variação dos parâmetros biofísicos coletados em campo em função da idade, DAP (a), H (b), Ht (c), Hc (d), Rc (e) e D (f).
Table 1 – Correlation coefficients (r) between dendrometric variables and $\rho_s$.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Correlation coefficients: significant for $p&lt;0.05$ highlighted (N=29)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho_1$</td>
</tr>
<tr>
<td>22 April $DAP$</td>
<td>0.03</td>
</tr>
<tr>
<td>22 April $H$</td>
<td>0.06</td>
</tr>
<tr>
<td>22 April $Ht$</td>
<td>0.12</td>
</tr>
<tr>
<td>22 April $HC$</td>
<td>-0.28</td>
</tr>
<tr>
<td>22 April $Rc$</td>
<td>0.07</td>
</tr>
<tr>
<td>22 April Densitometer</td>
<td>-0.01</td>
</tr>
<tr>
<td>24 May $DAP$</td>
<td>-0.03</td>
</tr>
<tr>
<td>24 May $H$</td>
<td>-0.07</td>
</tr>
<tr>
<td>24 May $Ht$</td>
<td>-0.15</td>
</tr>
<tr>
<td>24 May $HC$</td>
<td>0.37</td>
</tr>
<tr>
<td>24 May $Rc$</td>
<td>-0.11</td>
</tr>
<tr>
<td>24 May Densitometer</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Figure 4 – Thematic maps referring to the spatial distribution of dendrometric and structural parameters.

Figura 4 – Mapas temáticos relativos à distribuição espacial de parâmetros dendrométricos e estruturais.

To be continued...

Continua...

Figure 4 – Continued...

Figura 4 – Continua...

To be continued...

Continua...
Figure 4 – Continued...

**Spatial distribution of crown height (m)**

(g) $H_c = -3.698 + 0.2368 \times B4$

$h^2 = 0.60$

(h) $H_c = -1.3299 + 0.1604 \times B4$

$h^2 = 0.50$

**Spatial distribution of crown radius (m)**

(i) $R_c = 0.265 + 0.0268 \times B4$

$h^2 = 0.70$

(j) $R_c = 0.446 + 0.0219 \times B4$

$h^2 = 0.62$

To be continued...

Continua...
3.1 Results of regression models at different dates

A visual analysis of the thematic maps generated with data from the two dates reveals they were very similar, indicating consistency in the correlations established, despite the relatively small number of samples (20 points) being used for establishing the models. Regardless of the similarity, the proportion test revealed that the image acquired on 22/April/2009 had greater model fitting ‘capability’. Table 2 data show that, for all parameters, the value of the test statistic (calculated z-value) was higher than the critical value (critical z-value = 1.645). Therefore, for a 5% significance level, \( H_0 \) is rejected, in other words, the proportion of times the \( R^2 \) values associated to the April image were higher than the \( R^2 \) values associated to the May image is not equal to 50%. This means that the models fitted for the two images (dates) differed in their fitting capability, with the image from 22/April/2009 being significantly better.

Following a comparative assessment of the fitted models, an assessment was made of their predictive capability, using the 9 points that had been left unused.

The computed average percentage differences between MSPE and MSE (Table 3) indicated that they...
Structural characterization of canopies ...

Table 2 – Calculated values for the z-test.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Calculated z-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAP</td>
<td>97.40</td>
</tr>
<tr>
<td>H</td>
<td>98.80</td>
</tr>
<tr>
<td>Ht</td>
<td>99.04</td>
</tr>
<tr>
<td>Hc</td>
<td>99.54</td>
</tr>
<tr>
<td>Rc</td>
<td>99.36</td>
</tr>
<tr>
<td>D</td>
<td>95.70</td>
</tr>
</tbody>
</table>

were relatively similar and, consequently, the fitted models (for the two dates and each variable) had good predictive capability. For the variables H, Ht and Hc, the similarity of MSPR and MSE values was greater.

Overall, it was observed in this empirical approach that the selection/adoption, spontaneous or not, of a specific date for collecting orbital data would entail establishment of specific models, with greater or lesser fit capability, and yet with similar prediction capabilities to estimate structural variables of canopies.

4 CONCLUSIONS

The most significant correlations between radiometric and biophysical data in this study were observed for the near-infrared region (Band 4, TM/Landsat-5 sensor).

The spatial distribution of dendrometric data and canopy density (D) was consistent, at least visually, with the age distribution across the stands, indicating that even empirically it would be possible to estimate dendrometric data and D consistently, based on radiometric orbital data.

Despite the consistency, differences found in estimates (two dates) reinforce the fragility of this type of approach, which is often assumed to be discrete and absolute. Selecting an image acquisition date from which to extract radiometric data for model definition is a critical aspect of model performance, since in this study a significant performance difference was identified between them, depending on which date data were acquired.

5 REFERENCES


