Geospatial Estimation of above Ground Forest Biomass in the Sierra Madre Occidental in the State of Durango, Mexico

Pablito M. López-Serrano, Carlos A. López Sánchez, Raúl Solís-Moreno and José J. Corral-Rivas

1 Doctorado Institucional en Ciencias Agropecuarias y Forestales, Universidad Juárez del Estado de Durango, Boulevard Guadiana 501, Fraccionamiento Ciudad Universitaria, 34120 Durango, Mexico; pmslopez@gmail.com
2 Instituto de Silvicultura e Industria de la Madera, Universidad Juárez del Estado de Durango, Boulevard Guadiana 501, Fraccionamiento Ciudad Universitaria, 34120 Durango, Mexico; jcorral@ujed.mx
3 Facultad de Ciencias Forestales, Universidad Juárez del Estado de Durango, Río Papaloapan, Valle del Sur, 34120 Durango, Mexico; rsolis@ujed.mx
* Correspondence: calopez@ujed.mx; Tel.: +52-618-8251886
† These authors contributed equally to this work.

Abstract: Combined use of new geospatial techniques and non-parametric multivariate statistical methods enables monitoring and quantification of the biomass of large areas of forest ecosystems with acceptable reliability. The main objective of the present study was to estimate the aboveground forest biomass (AGB) in the Sierra Madre Occidental (SMO) in the state of Durango, Mexico, using the M5 model tree (M5P) technique and the analysis of medium-resolution satellite-based multi-spectral data, and field data collected from a network of 201 permanent forest growth and soil research sites (SPIFyS). Research plots were installed by systematic sampling throughout the study area in 2011. The digital levels of the images were converted to apparent reflectance (ToA) and surface reflectance (SR). The M5P technique that constructs tree-based piecewise linear models was used. The fitted model with SR and tree abundance by species group as predictive variables (ASG) explained 73% of the observed AGB variance (the root mean squared error (RMSE) = 39.40 Mg ha\(^{-1}\)). The variables that best discriminated the AGB, in order of decreasing importance, were the normalized difference vegetation index (NDVI), tree abundance of other broadleaves species (OB), Band 4 of Landsat 5 TM (Thematic Mapper) satellite and tree abundance of pines (Pinus). The results demonstrate the potential usefulness of the M5P method for estimating AGB based in the surface reflectance values (SR).

Keywords: M5P; remote sensing; Landsat; Sierra Madre Occidental

1. Introduction

The Sierra Madre Occidental (SMO) mountain range is of great ecological interest because of its environmental heterogeneity, which is attributed to the broad physiographical and climatic diversity in the area [1]. Moreover, the SMO is home to pine and oak species that are economically important in ecosystems in Mexico and other parts of the world [2]. The SMO crosses several states in western Mexico, including the state of Durango (the SMO occupies 71.5% of the surface area of the state). The state of Durango generates between 25% and 30% of the national timber production, producing a total of 1.5 million m\(^3\) of roundwood per year, and boasts forest reserves that are important sources of environmental services [3]. Studies that attempt to estimate forest biomass in this type of ecosystem...
are expensive due to its large coverage and difficult access for direct estimation of biomass. Thus, the emergence of geospatial techniques is becoming increasingly relevant for estimating and monitoring forest biomass in short periods of time because of its low cost and acceptable accuracy [4–7]. Because of the macrospatial scale and high heterogeneity of these ecosystems, the quantitative data obtained often do not comply with the underlying assumptions of simple statistical analysis (homogeneity and normality of distribution), so other techniques such as logistic regression and non-parametric classification methods are often applied [8–10]. The M5 model tree (M5P) technique is a reconstruction of M5 algorithm for inducing trees of regression models [11]. M5P is used for numeric prediction and at each leaf it stores a linear regression model that predicts the class value of instances that reach the leaf. To determine which attribute is the best to split the portion of the training data that reaches a particular node, the splitting criterion is used. The standard deviation of the training class is treated as a measure of the error at that node and each attribute at that node is tested by calculating the expected reduction in error. The attribute that is chosen for splitting maximizes the expected error reduction at that node. The main objective of the present study was to estimate the aboveground forest biomass (AGB) in the SMO in the state of Durango, Mexico, using the M5P technique and the analysis of medium-resolution satellite-based multi-spectral data, and field data collected from a network of 201 permanent forest growth and soil research sites (SPIFYs).

2. Material and Methods

2.1. Study Area

The study area is a mountainous zone in the state of Durango (Mexico) that forms part of the Sierra Madre Occidental (Figure 1).
The area occupied by the state of Durango represents 6.3% of the land in Mexico. The total area covered by the state is 12.3 million ha, of which 9.1 million ha (74.35% of the land in the state) is forestland managed by 11 Regional Forest Management Units (UMAFORES). A large part of the forestland (4.9 million ha) is occupied by temperate forest and is subjected to precipitation levels of between 800 and 1200 mm per year, with frost occurring in winter as a result of the combination of low temperatures and humid winds from the Pacific Ocean; a smaller area of the land (0.5 million), affected by warmer climate, is occupied by forest classified as rainforest [12]. The mean elevation in this zone is 2650 m above sea level. These forests have rich biodiversity and include at least 27 coniferous tree species (of which 20 are Pinus species) and 43 species of Quercus; the predominant forest stands comprise pines and oaks, often mixed with Arbutus and Juniperus, among other tree species [13]. These unique forests are irregular and have been subject to selective harvesting for almost a century to provide a mix of services to local communities. This irregularity refers to the spatial arrangement of trees (vertical and horizontal irregularity) and the variation in the age structure of trees and stands. This structure is the result of the management history, which has depended on land ownership, as well as the economic and social changes that have taken place in the state, and also natural conditions [14].

2.2. Field Data

The dasometric data were obtained from 201 permanent forest growth and soil monitoring plots (SPIFyS) in the SMO in the state of Durango. The plots were installed during the winter of 2011 using the protocol developed by [15]. The data of these permanent sample plots are used to monitor the growth and yield of Durango’s forests. The plots cover the main forest types and the current diameter distributions of commercial forests in Durango. The plots are 50 × 50 m in size (distance was corrected by the slope) and are distributed by systematic sampling (with some exceptions), with a variable grid ranging from 3 to 5 kilometers, depending on the size of the “Ejidos”. Ejidos are communal groups that live in rural areas and whose lands are managed with some level of governmental control. The sampling plots are intended to be re-measured at five-year intervals. Among other variables, tag number, species code, breast height diameter (measured in cm at 1.3 m above ground level), total tree height (m), height to the live crown (m), azimuth (°) and radius (m) from the center of the plot of all trees equal or larger than 7.5 centimeters (cm) in diameter were recorded. The database used here includes measurement data from 31,979 trees.

The aboveground biomass in each of the SPIFyS plots was estimated using specific allometric equations developed by [16] for the same study area. Depending on the species, the goodness of fit statistics ranged between 0.82 and 0.97 of the coefficient of determination ($R^2$) and the root mean square error (RMSE) between 22.68 and 133.68 kg.

The main descriptive statistics for the total aboveground biomass per hectare in the study sites are summarized in Table 1.

Table 1. Descriptive statistics of the total aboveground biomass per hectare in the 201 permanent forest growth and soil monitoring plots (SPIFyS).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum Value</th>
<th>Maximum Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of stems per ha</td>
<td>645</td>
<td>271.84</td>
<td>224</td>
<td>2264</td>
</tr>
<tr>
<td>Stand basal area (m$^2$.ha$^{-1}$)</td>
<td>23.44</td>
<td>8.06</td>
<td>8.21</td>
<td>54.83</td>
</tr>
<tr>
<td>Dominant height (m)</td>
<td>17.47</td>
<td>5.08</td>
<td>6.86</td>
<td>30.60</td>
</tr>
<tr>
<td>Stand biomass (Mg. ha$^{-1}$)</td>
<td>141.64</td>
<td>75.01</td>
<td>27.73</td>
<td>469.42</td>
</tr>
</tbody>
</table>

2.3. Tree Abundance by Species Group (ASG)

The tree abundance (number of trees per group of species per plot) was estimated for posterior analysis in this study. A total of seventy-two different tree species were grouped in four groups of species as they present similar growth patterns: (P) Pinus species (16); (OC) other conifers species (12); (Q) oaks species (26); and (OB) other broadleaves species (18).
2.4. Source of Spectral Data

The data used for the study were obtained from six Landsat 5 TM (Thematic Mapper) satellite images captured between March and May 2011 and covering all of the SMO within the state of Durango (path/row: 30/44, 31/42, 31/43, 31/44, 32/42 and 32/43) [17]. This satellite platform, of medium spatial resolution, operates in seven bands of the electromagnetic spectrum: blue (bandwidth 0.45–0.52 µm), green (bandwidth 0.52–0.60 µm), red (0.63–0.69 µm), near infrared (0.78–0.89 µm), mid infrared (1.55–1.75 µm) and far infrared (2.08–2.35 µm). These bandwidths correspond, respectively, to bands 1, 2, 3, 4, 5 and 7 of the Landsat TM5 satellite [18]. Band 6, designed for the thermal mapping and soil moisture, was not considered because of its lower (120 m) spatial resolution.

The satellite images were digitally pre-processed by radiometric correction techniques, according to the procedures suggested by [19,20]. The images are produced by USGS with a rectification using a cubic convolution geometric correction for discrete data (level L1T), with a root mean square error (RMSE) of less than 1 pixel, thus making them suitable for digital image processing [21]. The digital levels (DLs) were converted to radiance values to generate images that were calibrated with the minimal radiance (Lmin) and maximal radiance (Lmax) values for each band of the sensor [22]. The radiance was subsequently converted to apparent reflectance (Top of Atmosphere (ToA)) with the aim of converting the original values of each image into standard physical variables that are comparable over time for the same sensor [19]. This process was carried out with IDRISI® Selva software [23] and the ATMOS algorithm, which fits the radiometric effect on considering the solar elevation, yielding an image with reflectance values (0–1).

Furthermore, the same images were downloaded from the National Landsat Archive Processing System (NLAPS), corresponding to the product Landsat 4–5 Thematic Mapper level 1 of reflectance on surfaces (SR), radiometrically and atmospherically corrected, and processed through the Standard Landsat Product Generation System (LPGS) using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm [17].

Bands 1, 2, 3, 4, 5 and 7 (Band 1 to Band 7) of Landsat TM5 were used; Band 6 was not used, because of its thermal characteristics [17]. The Normalized Difference Vegetation Index (NDVI) was also calculated, with the aim of compensating the factors that influence the images in relation to biomass estimation, such as the illumination conditions, the slope and the orientation of the surface. The use of NDVI calculated for ToA and SR, as a predictor variable to model AGB has been successfully reported in previous studies [24–26].

\[
NDVI = \frac{(NIR - R)}{(NIR + R)}
\]  

where \(NIR\) is the spectral band in the near infrared region (Band 4) and \(R\) the band in the red region (Band 3).

2.5. Integration of Data Files

Once the images were obtained by the previously mentioned processes (ToA and SR spectral bands), a mosaic was constructed with six of the scenes covering the SMO. Posterior geolocation of the SPIFyS in the mosaic enabled extraction of the information at the pixel level by bilinear interpolation. ArcGIS 10® software [27] was used for this extraction. ToA and SR values were integrated in a database together with the extracted total aboveground biomass (Mg ha\(^{-1}\)) as inputs for the model.

2.6. Fitted Model

We used a machine learning technique to estimate the AGB at stand level. M5P technique combines a conventional decision tree with the possibility of linear regression functions at the nodes. First, a decision-tree induction algorithm is used to build a tree, but instead of maximizing the information gain at each inner node, a splitting criterion is used that minimizes the intra-subset
variation in the class values down each branch. The splitting procedure in M5P stops if the class values of all instances that reach a node vary very slightly, or only a few instances remain. Second, the tree is pruned back from each leaf. When pruning, an inner node is turned into a leaf with a regression plane. Third, to avoid sharp discontinuities between the subtrees, a smoothing procedure is applied that combines the leaf model prediction with each node along the path back to the root, smoothing it at each of these nodes by combining it with the value predicted by the linear model for that node. Techniques devised by [11] for their classification and regression trees system are adapted in order to deal with enumerated attributes and missing values. All enumerated attributes are turned into binary variables so that all splits in M5P are binary. As to missing values, M5P uses a technique called “surrogate splitting” that finds another attribute to split on in place of the original one and uses it instead [11,28,29].

In this study, in a first stage, the six spectral bands of the Landsat 5 TM sensor (1, 2, 3, 4, 5 and 7) and the NDVI were analyzed with the algorithms ToA and SR to estimate AGB. In a second stage, its spectral variables (SR) were evaluated with variables that incorporate aspects of forest structure (ASG). All analyses were performed with M5P technique implemented into the WEKA open source software [30].

To compare the performance of the models, the coefficient of determination ($R^2$), the root mean squared error (RMSE) and the root relative squared error (RRSE) were used as goodness-of-fit criteria for evaluating model performance and were expressed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - p}}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}}$$

where, $y_i$, $\hat{y}_i$ and $\bar{y}_i$ are the observed, estimated and mean values of AGB, respectively; $n$ is the total number of observations used to fit the model; and $p$ is the number of model parameters.

The selected model was applied for mapping AGB in the SMO area using ArcGIS 10® software [27].

3. Results

The decision tree generated by the M5P technique for ToA, SR and SR with ASG variables were implemented in WEKA software, using the pixel level values extracted from the images of the 201 SPIFyS plots is shown in Figure 2. In accordance with the hierarchical structure of the decision trees, the following variables that best discriminated or predicted the AGB, in order of decreasing importance were, for ToA: Band 7, Band 3, Band 1, NDVI and Band 5; for SR: NDVI, Band 1 and Band 7; and for SR with ASG: NDVI, OB, Band 4 and tree abundance of pines. Categorization of the trees continued following the path determined by the responses to the questions at the internal nodes, until reaching a terminal node, where the predetermined label will be that assigned to the classification pattern—in this case, the pixel values for AGB estimation. The Table 2 show the goodness-of-fit statistics derived from the M5P technique with ToA values explained 54% ($R^2$) of the observed variability in the AGB of the 201 research plots, with a
RMSE of 50.47 Mg·ha\(^{-1}\). SR model explained 69% (RMSE = 42.17 Mg·ha\(^{-1}\)), and when including the ASG variables the explanation of the variance increases to 73% (RMSE = 39.40 Mg·ha\(^{-1}\)).

**Figure 2.** Decision tree obtained using the M5P technique with ToA (upper), SR (middle) and SR with ASG variables (bottom). (Band 1 to 7): of Landsat 5 TM (Thematic Mapper) satellite, (NDVI): normalized difference vegetation index, (OB) tree abundance of other broadleaves species, and (Pinus) tree abundance of pines.
Table 2. Summary of the goodness-of-fit statistics for estimation of the AGB.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>ToA</th>
<th>SR</th>
<th>SR with ASG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.54</td>
<td>0.69</td>
<td>0.73</td>
</tr>
<tr>
<td>RMSE</td>
<td>50.47</td>
<td>42.17</td>
<td>39.40</td>
</tr>
<tr>
<td>RRSE</td>
<td>67.45</td>
<td>56.36</td>
<td>52.66</td>
</tr>
</tbody>
</table>

Graphical analysis of the residual values and the observed values plotted against the predicted values of AGB did not reveal any important problems in relation to heterogeneity of the variance or lack of normal distribution of the residuals, with the exception of a slight trend of underestimation for high AGB (Figure 3).

Figure 3. Graphs showing the distribution of the residuals and of the observed AGB values with ToA (upper), SR (middle) and SR with ASG variables (bottom).
The spatial distribution of the estimated AGB (Mg·ha⁻¹) in the SMO area obtained by the application of the classification rules included in the regression tree model (M5P) for SR variables is shown in Figure 4. The lighter color pixels represent the lowest amounts of AGB, below 75 Mg·ha⁻¹, whereas the dark green pixels represent the largest amounts of AGB, which consistently correspond to the most dense areas of temperate forest. Calculated mean amount of AGB for the study area was around 106 Mg·ha⁻¹.

![Figure 4. Spatial distribution of the total AGB in the SMO, state of Durango, Mexico.](image)

The total AGB content estimations from the MP5 technique for SR variables for the analyzed forest management units are shown in Table 3. The highest mean value of AGB was observed in the UMAFOR 1006 (Municipally of San Dimas) with 148.98 Mg·ha⁻¹ and a total estimation of 64,033,008.59 Mg. This zone encompasses the largest area of forestland and therefore the largest amount of AGB. On the other hand, the lowest amount of AGB was observed in the UMAFOR 1001 with a mean estimate of 78.66 Mg·ha⁻¹, making it the forest region with the lowest density out of the eleven forest management units considered in this study.
Table 3. Estimation of AGB for the regional forest management units in the SMO, state of Durango, Mexico.

<table>
<thead>
<tr>
<th>UMAFOR</th>
<th>Mean AGB (Mg·ha⁻¹)</th>
<th>Surface Area (ha)</th>
<th>Total AGB (Mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td>78.66</td>
<td>423,990.00</td>
<td>33,350,319.02</td>
</tr>
<tr>
<td>1002</td>
<td>85.58</td>
<td>351,498.00</td>
<td>30,079,977.18</td>
</tr>
<tr>
<td>1003</td>
<td>107.10</td>
<td>126,054.00</td>
<td>13,500,870.12</td>
</tr>
<tr>
<td>1004</td>
<td>99.54</td>
<td>318,104.00</td>
<td>31,663,478.70</td>
</tr>
<tr>
<td>1005</td>
<td>125.38</td>
<td>424,753.00</td>
<td>53,256,210.80</td>
</tr>
<tr>
<td>1006</td>
<td>148.98</td>
<td>429,806.00</td>
<td>64,033,008.59</td>
</tr>
<tr>
<td>1007</td>
<td>88.29</td>
<td>253,619.00</td>
<td>22,393,159.18</td>
</tr>
<tr>
<td>1008</td>
<td>111.12</td>
<td>373,308.00</td>
<td>41,482,686.80</td>
</tr>
<tr>
<td>1009</td>
<td>120.90</td>
<td>162,075.00</td>
<td>19,594,636.15</td>
</tr>
<tr>
<td>1010</td>
<td>111.49</td>
<td>358,944.00</td>
<td>40,017,365.50</td>
</tr>
<tr>
<td>1011</td>
<td>84.97</td>
<td>262,488.00</td>
<td>22,303,475.59</td>
</tr>
<tr>
<td>TOTAL</td>
<td>105.64</td>
<td>3,484,639.00</td>
<td>283,143,798.25</td>
</tr>
</tbody>
</table>

4. Discussion

The results of the present study demonstrate that the data acquired by a medium spatial resolution (Landsat) sensor are potentially useful for estimating AGB in structurally complex forests, such as those in the SMO in the state of Durango (Mexico), with satisfactory results and low cost. The deterministic predictors were the bands belonging to the blue, green and most red, near and mid infrared spectral regions. This finding was similar to that reported by [31], who demonstrated that reflectance in the red and near infrared regions yielded good predictions for AGB estimation in forest zones of the Yellowstone National Park, USA. In the present study, the model tended to underestimate AGB values above approximately 250 Mg·ha⁻¹. This might possibly be due to the saturation of NDVI, which is the most influential variable in predicting biomass for high values. In this sense, several studies have similarly found that the NDVI loses its sensitivity to dense vegetation because of the saturation in red and near infrared wavelength in measuring and monitoring plant growth, vegetation cover and biomass production from satellite data [32–34]. Models fitted ToA, SR, and SR with ASG, respectively, showed an increasing capacity to overcome these NDVI saturation problems.

Furthermore, [35] concluded that the vegetation indices or individual bands, which include one or more bands in the infrared spectrum, provide satisfactory descriptions of zones occupied by conifer or broadleaf species. Moreover, in the case of SR, which was better than ToA, the contribution of Band 1 in two terminal nodes of the M5P model is associated with the structural variability of the canopy [36]. Günlü et al. [37] found that the reflectance from Landsat TM satellite Band 1 was the best predictor of AGB ($R^2 = 0.465$, RMSE = 91,836 t·ha⁻¹), given the structural conditions of the canopy and understory, as the reflectance from this band increased as the AGB increased. In the present study, the result of the M5P analysis with SR spectral bands ($R^2 = 0.69$, RMSE = 42.17 Mg·ha⁻¹) was higher to that reported by Houghton et al. [38], who analyzed data from the MODIS sensor (resolution, 500 m) and forest inventory data using the non-parametric Random Forest ($R^2 = 0.61$) method to map forest biomass in Russia.

In a recent study, Tian et al. [39] used the non-parametric $k$-nearest neighbours ($k$-NN) technique to produce an optimized model ($R^2 = 0.59$, RMSE = 24.92 ton·ha⁻¹) from Landsat-TM images of a sample of 133 plots, with topographic correction based on sun-canopy-sensor (SCS + C). Likewise, Hall, [30] used Landsat 5 (TM) images rectified by SCS+C radiometric correction and compared the performance of the $k$-NN method and support vector machine (SVM) method for estimating AGB. They found that $k$-NN performed better ($R^2 = 0.54$; RMSE = 26.62 ton·ha⁻¹) than SVM ($R^2 = 0.51$; RMSE = 27.45 ton·ha⁻¹).

In general, most previous studies report significant relationships between AGB and the reflectance values yielded by each sensor. The reliability was within the range reported in diverse research...
studies that estimate AGB from medium resolution spectral data, from Landsat and SPOT, which often yield $R^2$ values between 0.50 and 0.70 with absolute errors of the estimates of between 30 and 60 Mg·ha$^{-1}$ [5,40–47]. The present study also shows that incorporation of spectral data and tree abundance estimated by species group in mixed and uneven-aged forests (SR with ASG), such as the SMO, can increase the level of estimation of the AGB ($R^2 = 0.73; \text{RMSE} = 39.40 \text{Mg·ha}^{-1}$). In this sense, previous studies have reported significant variations in forest biomass estimation between different ecological zones, tree species, ages, density and management types [48–50].

In other studies [51–53], several authors have concluded that the spectral data derived after atmospheric and topographic correction may improve the accuracy of the biomass estimation, irrespective of the statistical method used. As the areas being monitored are mountainous zones, the quality of the data is negatively affected by the reflectance between sunny and shaded slopes. Interactive parameter fitting in the topographical correction methods may improve the quality of the spectral data and of the AGB estimates [52,53].

5. Conclusions

In the present study, we estimated the AGB in the SMO in the state of Durango, Mexico, using the MSP technique and the analysis of medium-resolution satellite-based multi-spectral data, and field data collected from a network of 201 SPIFyS.

The findings show that the MSP method is potentially useful for estimating forest biomass. Data from the infrared channel of the Landsat TM5 sensor proved best for discriminating or predicting AGB.

The surface reflectance values (SR) in comparison with atmospheric correction from the sensor (ToA), was best for the estimation of AGB.

The results of this study indicate that performing atmospheric corrections and considering variables related to forest structure (SR with ASG variables) can help to solve problems of saturation of NDVI for high values of biomass.

Acknowledgments: This research was supported by SEP-PROMEP (project: Seguimiento y Evaluación de Sitios Permanentes de Investigación Forestal y el Impacto Socio-económico del Manejo Forestal en el Norte de México).

Author Contributions: Pablito M. López-Serrano and Carlos A. López-Sánchez conceived, designed and performed the experiments and wrote the manuscript. Raúl Solís-Moreno and José J. Corral-Rivas analyzed the data and revised the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References


© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).