



## **Landsat ETM+ imaging for the estimation of the forest density in the southern region of the State of Mexico**

Miguel Acosta Mireles<sup>1\*</sup>

Ramiro Pérez Miranda<sup>2</sup>

Martín Enrique Romero Sánchez<sup>2</sup>

Antonio González Hernández<sup>2</sup>

Luis Martínez Ángel<sup>3</sup>

<sup>1</sup> Campo Experimental Valle de México- CIR-Centro. INIFAP. México.

<sup>2</sup> Centro Nacional de Investigación Disciplinarias en Conservación y Mejoramiento de Ecosistemas Forestales. INIFAP. México.

<sup>3</sup> Instituto Tecnológico Superior de Venustiano Carranza (ITSVC). México.

\*Autor por correspondencia: Miguel Acosta Mireles, Correo-e: [acostamm@colpos.mx](mailto:acostamm@colpos.mx)



**Abstract:**

In the context of the mechanisms of climatic change mitigation, constant forest monitoring is important because forests provide crucial information. The estimation of forest stand attributes based on satellite imagery data combined with forest inventory data allows producing accurate information on forest structure, at a relative and accessible cost. However, there is still a need to use models that enable the construction of a valid relationship between remote sensing and field data. Therefore, this study aims to estimate forest attributes such as basal area (AB), volume (V) and aboveground biomass (B) by analyzing and using the relationship between spectral information, from Landsat ETM+ imagery, and tree measurements. Data from plots collected in 2010 by the National Forest and Soils Inventory (INFyS) were used as primary source of information. Results suggest that the best models to estimate AB, V and B were those that used the near infrared band (band 5) as the independent variable. Results also indicated that adjusted regression models shown statistical bases to estimate AB, V and B in a precise manner. All regression models were highly significant (5 %) with determination coefficients ( $R^2_{adj}$ ) higher than 0.47.

**Key words:** aboveground biomass, forest structure, Landsat, regression models, forest parameters, remote sensors.



## Introduction

Forest ecosystems capture, store and release carbon as a result of photosynthetic and respiratory processes and dry matter degradation (Razo *et al.*, 2013), which contributes directly to the abatement of the effects of climate change carbon sequestration in their biomass. Therefore, within the context of climate change abatement, forest monitoring is important because forests provide crucial information on the effects of the process (Sinha *et al.*, 2015).

Aboveground biomass and other forest parameters (e.g. volume, basal area, carbon and leaf area index) are estimated using the direct and indirect methods. The former consists of a destructive sampling; however, it is very slow and costly, which makes its large scale application complicated. The latter is based on: i) statistical techniques (allometric equations and biomass expansion factor) (Ayala *et al.*, 2002) relating variables that can be easily measured (normal diameter and total height) in forest inventories with variables that are difficult to measure (e.g. volume and biomass) (Návar, 2009; Aquino *et al.*, 2015); and ii) models obtained from combinations of data derived from the forest inventory and from remote sensing (Labrecque *et al.*, 2006; Hall *et al.*, 2006).

The enhancement of the capacities of the various types of remote sensors provides the opportunity to develop more efficient analysis techniques (Torres *et al.*, 2016), which in turn favors the obtainment of more consistent results in the assessment and monitoring of forest resources. Remote sensors have comparative advantages over the conventional inventory: optimization in time and financial supports, access to inaccessible areas, and the execution of inventories in large forest surface areas (Hou *et al.* 2011; Sinha *et al.*, 2015; Timothy *et al.*, 2016). For this reason, the estimation of the forest biomass based on this technology is of great interest to achieve a

sustainable forest management and to generate environmental policies (Latifi *et al.*, 2015).

The use of satellite imaging as auxiliary variables has made it possible to monitor the forest resources constantly and to accurately estimate various attributes of this kind (Dong *et al.*, 2003; Valdez *et al.*, 2006; Muñoz *et al.*, 2014). However, the relationship between data from remote sensing and field data must be further explored in order to estimate the forest parameters. The objective of the study was to develop regression models for the estimation of the basal area (BA), volume (V) and biomass (B) with auxiliary variables of spectral data and vegetation indices obtained through Landsat 7 ETM+ imaging.

## Materials and Methods

### Study area

The study area is located in the southeastern end of the State of Mexico and comprises a surface area of 409 936 ha; between the coordinates 19°15'00" N and 100°37'00" W, and 18°22'00" N and 99°45'00" W (Figure 1), within the hydrological region of the *Balsas* river. Its climates are semiwarm ((A)C(w) and A(C)w) and warm, with a mean annual temperature of 17.3 °C, and a mean annual precipitation of 939 mm (García-Conabio, 1998).

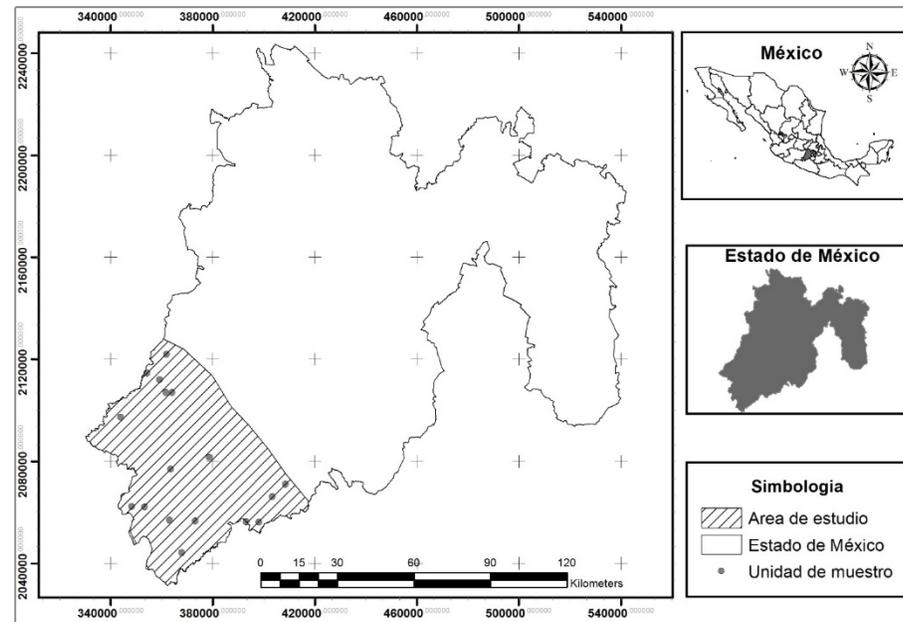
The main mighty rivers in the area are *Temascaltepec*, *Sultepec*, *Topilar*, *San Pedro*, *Amacuzac* and *Cutzamala* (GEM, 2007). As for the orography, the region includes the *Sierra Temascaltepec* mountain systems, whose morphology is rough and uneven, with narrow valleys, gullies and ravines (GEM, 2007).

The prevalent geology consists of carbonated and vulcanosedimentary rocks of the Higher Jurassic and lower Cretaceous periods, vulcanosedimentary rocks of the

Higher Triassic and Lower Jurassic, and intrusive felsic rocks of the Tertiary period (GEM, 2007).

The soils are mainly Regosols, Cambisols, Andosols and Phaeozems (Inifap, 1995).

The vegetation corresponds to a low deciduous forest; there are also *Quercus* spp. and *Pinus* spp. forests, induced grasslands and rainfed agriculture (Inegi, 2013).



**Figure 1.** Location of the study area in the State of Mexico.

## Field data

17 clusters of the 2010 Forest and Soils Inventory of the State of Mexico (Probosque, 2010) corresponding to low deciduous forests of the southern extreme of the entity (Figure 1) were used. Before calculating the variables, the database was cleared, and the atypical values for the normal diameter (D) and total height (H) were deleted.

Each cluster consists of four rectangular sampling sites with a surface area of 400 m<sup>2</sup> (Conafor, 2011). Information regarding the dasometric variables is available for each

sampling site. The estimations of the variables BA (m<sup>2</sup>), V (m<sup>3</sup>) and B (kg) were registered at tree level. The BA was estimated based on the 0.7854\*DN<sup>2</sup> structure, in which D is the normal diameter and 0.7854 is the constant resulting from the  $\pi/4$  ratio.

The volume was estimated using two equations developed for the low deciduous forest: 1) the equation of the National Forest and Soils Inventory (1973), and 2) the equation derived from a manifestation of environmental impact in the regional modality of the advanced management programs for the forests of eight *ejidos* in the state of *Morelos* (Sinat, 2007) (Table 1). The general equation designed by Torres and Guevara (2002) was used to estimate the total biomass in the low forest (Table 2).

**Table 1.** Equation to calculate the volume.

Equation	Model	$\beta_0$	$\beta_1$	$\beta_2$	
1	$V = e^{\beta_0} D^{\beta_1} H^{\beta_2}$	-0.77785	1.872175	0.815238	1)
2		-9.3156	2.38434	0.1666	

Where:

$V$  = Total volume (kg)

$\beta_i$  = Value of the parameters

**Table 2.** Total biomass equation for the low deciduous forest.

Equation	Model	$\beta_0$	$\beta_1$	
3	$B = \beta_0 + \beta_1 \times (V)$	12.225	313.036	2)

Where:

$B$  = Total biomass (kg)

$\beta_i$  = Value of the parameters

For each of the forest variables (BA, V and B), the addition was calculated per site and per cluster. In order to extrapolate the values per hectare, the ratio of means method was applied to each variable (Šmelko and Merganič, 2008):

$$\check{R} = \frac{Y}{X} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i} \quad 3)'766 \text{ Where:}$$

$\check{R}$  = Variable of interest expressed in has

$Y_i$  = Total value of the variable in all the 400 m<sup>2</sup> sites

$X_i$  = Total sampled surface area in  $i$  sites.

## Spectral variables

A Landsat 7 ETM+ was obtained through the United States Geological Survey (USGS, 2015), with a spatial resolution of 30 x 30 m (900 m<sup>2</sup>), at LT 1 and corrected by the same processing system of the same American service (Landsat Ecosystem Disturbance Adaptive Processing System: LEDAPS), which carries out geometric and radiometric corrections (Masek *et al.*, 2006). The date on which the image was taken was July 30, 2010. The mean values of the pixels located within a 1 ha mask and corresponding to the size of a duly georeferenced sampling cluster were estimated (Hall *et al.*, 2006; Muñoz *et al.*, 2014). Three vegetation indexes (VI) were estimated using mathematical transformations, as these reflect their status and allow modeling the forest parameters accurately (ERDAS, 2011; Wijaya *et al.*, 2010; Poulain *et al.*, 2010; Muñoz *et al.*, 2014).

$$NDVI = \frac{NIR-R}{NIR+R} \quad 4)$$

$$GNDVI = \frac{NIR-G}{NIR+G} \quad 5)$$

$$DVI = NIR - R \quad 6)$$

Where:

*NDVI* = Normalized difference vegetation index

*GNDVI* = Green normalized vegetation index

*DVI* = Difference vegetation index

*NIR* = Near-infrared band

*R* = Red band

*G* = Green band

## **Analysis of the correlation between forest parameters and spectral data**

A Pearson's correlation analysis was carried out to assess the correlation coefficient, the  $R^2$  and the level of significance to a rejection value under ( $\alpha = 5\%$ ) in order to determine the degree of association between the parameters BA, V and B (dependent variable and/or response) and each of the spectral bands and the vegetation index (independent or predictive variable).

## Linear regression adjustment models

The variables that showed the highest correlation were used to build the linear regression models in order to estimate the BA, V and B. The models were fitted using the minimum ordinary squares (MOS) method with the SAS/ETS™ statistical package (SAS, 2008). The MOS method provides the best unbiased linear estimators.

The selected models were evaluated based on the accuracy and numeric precision of the adjustment statistics: determination coefficient ( $R^2_{adj}$ ), root of the mean square error (RMSE) and high significance in the parameters ( $\alpha = 0.05 \%$ ). The highest and lowest  $R^2_{adj}$  values corresponded, respectively, to a greater precision and accuracy of a model for estimating the basal area, volume and biomass. The best linear regression models were used to map and obtain the spatial distribution of the forest parameters of interest for each pixel in the image of the study area.

## Estimation of the forest parameters in terms of total inventories

In order to estimate the total inventory of the forest parameters in the study area, simple random sampling (traditional inventory) and the regression estimator (alternative inventory) were used. The latter includes an auxiliary measure (spectral bands and vegetation indices) that is highly correlated with the forest parameters (Valdez *et al.*, 2006; Ortiz *et al.*, 2015). Therefore, the spectral variables and vegetation indices with the highest degree of correlation with the forest parameters were handled as auxiliary variables in the calculation of the regression estimators as an alternative for updating the total inventory (Muñoz *et al.*, 2014; Ortiz *et al.*, 2015). For the calculations of the total inventory, a surface area of 1 000 has was used; this

made it possible to observe, compare and determine which method yielded the best estimation in terms of the total inventory (traditional inventory vs alternative inventory using remote sensing).

## Results and Discussion

### Analysis of the correlation between forest parameters and spectral data

The average volume calculated for the low deciduous forest was 30.26 and 22.08 m<sup>3</sup> ha<sup>-1</sup>, estimated with the V<sub>1</sub> and V<sub>2</sub> equations, respectively. Among the most contrasting results are those obtained by Probosque (2010), Del Ángel-Mobarak (2012) and Conafor (2012), with figures of 35.01, 29.58 and 23.72 m<sup>3</sup> ha<sup>-1</sup>, respectively, for the national volume in the low deciduous forest. Thus, the mean volume in this study is within the national range registered by other studies, which allows the use of either of the two equations for subsequent statistical procedures. The mean basal area is close to the value calculated by Del Ángel-Mobarak (2012), which is 4.74 m<sup>2</sup> ha<sup>-1</sup>; this result is similar to the one obtained by the present study —5.92 m<sup>2</sup> ha<sup>-1</sup>— for the aforementioned vegetation type. The mean value observed for the biomass in equations 1 and 2 ranged between 10.71 and 13.27 Mg ha<sup>-1</sup>.

The forest parameters (basal area, volume and biomass) showed a strong negative correlation with the spectral bands and a positive correlation with the vegetation indices. However, the highest correlation was with the mid-infrared (MIR) spectral band, with a value for R<sup>2</sup> ranging between -0.69 and -0.77, while the R<sup>2</sup> values for the normalized difference vegetation index (NDVI) ranged between 0.56 and 0.61, with a significance level of 5 % (Table 3).

**Table 3.** Pearson's correlation coefficients between spectral variables and the forest variables basal area (BA), volume (V) and biomass (B).

<b>Variables</b>	<b>Statistic</b>	<b>BA</b>	<b>V1</b>	<b>V2</b>	<b>B1</b>	<b>B2</b>
Band 1 (blue)	$R^2$	-0.46	-0.42	-0.38	-0.46	-0.43
	$\alpha = 5 \%$	0.0779	0.1123	0.1609	0.0819	0.1065
Band 2 (green)	$R^2$	-0.71	-0.66	-0.66	-0.72	-0.72
	$\alpha = 5 \%$	0.0028	0.0068	0.0067	0.0024	0.0023
Band 3 (red)	$R^2$	-0.70	-0.68	-0.70	-0.72	-0.73
	$\alpha = 5 \%$	0.003	0.0051	0.0033	0.0022	0.0016
Band 4 (near-infrared)	$R^2$	-0.41	-0.38	-0.38	-0.41	-0.41
	$\alpha = 5 \%$	0.1263	0.1514	0.162	0.1198	0.1262
Banda 5 (mid-infrared)	$R^2$	-0.72	-0.69	-0.74	-0.74	-0.77
	$\alpha = 5 \%$	0.0002	0.0002	0.0016	0.0014	0.0006
Banda 7 ( mid-infrared)	$R^2$	-0.64	-0.61	-0.67	-0.66	-0.70
	$\alpha = 5 \%$	0.0101	0.0144	0.0057	0.0068	0.0034
NDVI	$R^2$	0.58	0.56	0.58	0.59	0.61
	$\alpha = 5 \%$	0.0225	0.0293	0.0211	0.0187	0.0152
GNDVI	$R^2$	0.58	0.53	0.54	0.59	0.59
	$\alpha = 5 \%$	0.0226	0.0379	0.0344	0.0204	0.0188
DVI	$R^2$	-0.20	-0.18	-0.17	-0.20	-0.19
	$\alpha = 5 \%$	0.4666	0.5013	0.535	0.4616	0.4821

$BA =$  Basal area  $ha^{-1}$ ;  $V_i =$  Volume  $ha^{-1}$ ;  $B_i =$  Biomass  $ha^{-1}$

The mid-infrared (MIR) band was selected as independent variable, as it shows the highest correlation with the forest variables, which made it possible to have the best adjustment statistics in the assessed models. The negative correlation of the mid-infrared band is inversely proportional to forest density, which may be ascribed to the reduction of the albedo in high-density areas (Aguirre *et al.*, 2007; Aguirre *et al.*, 2009). Therefore, a significant increase in the MIR values suggests an increased amount of chlorophyll and, consequently, an increase in basal area, volume and biomass. The high associated correlations between the forest parameters (BA, V and B) and the MIR band made it possible to obtain more statistically consistent regression models.

## Linear regression models

The spatial distribution of the forest parameters (BA, V and B) of tropical forests is very complex; the topography is one of the main factors affecting the heterogeneity, which makes it difficult to develop an accurate model. However, in this study it was possible to obtain the most statistically consistent regression model for the estimation of the forest variables, which was built using the PROC MODEL procedure of the SAS/ETS™ statistical software package (SAS, 2008).

Table 4 shows the values of the fit statistics and the parameter estimators of the best regression models evaluated for the estimation of the forest variables. The intercept of the model was statistically different from zero ( $\beta_0$ ), and the rate of change in the slope of the mid-infrared variable ( $\beta_1$ ) contributed to estimate the variables BA, V and B, as the probabilities associated to the parameters are highly significant in the hypothesis test, at a significance level under 5 % ( $\alpha=0.05$ ).



**Table 4.** Fit and statistical values of the assessed models.

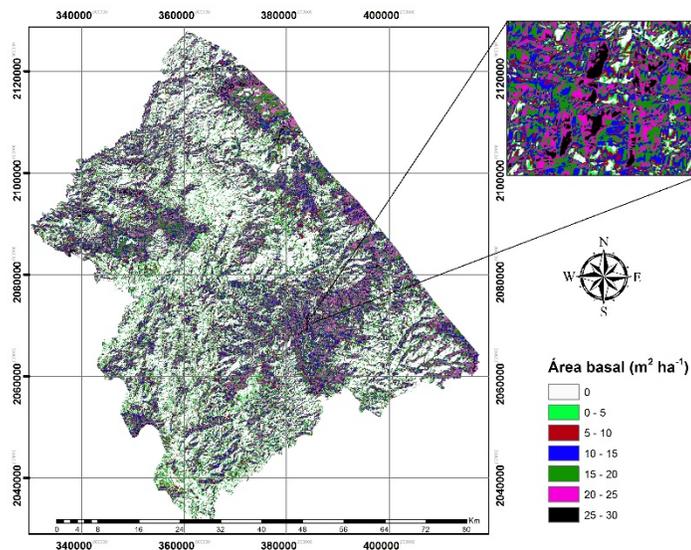
Model	Equation	$R^2_{adj}$	RMSE	$\rho$	$\Psi$	$\epsilon^t$	$\infty$
1	$AB_1 = \alpha_0 + \alpha_1 X_1$	0.52	3.30	$\alpha_0$	86.95	20.40	0.0008
				$\alpha_1$	-763.71	164.50	0.0004
2	$V_1 = \beta_0 + \beta_1 X_1$	0.47	17.92	$\beta_0$	452.51	108.12	0.0009
				$\beta_1$	-3946.44	871.68	0.0005
3	$V_2 = \beta_0 + \beta_1 X_1$	0.54	13.18	$\beta_0$	363.78	81.95	0.0006
				$\beta_1$	-3184.22	660.66	0.0003
4	$B_1 = \gamma_0 + \gamma_1 X_1$	0.55	7.25	$\gamma_0$	202.45	44.48	0.0005
				$\gamma_1$	-1777.03	358.64	0.0002
5	$B_2 = \gamma_0 + \gamma_1 X_1$	0.60	5.91	$\gamma_0$	174.68	37.88	0.0004
				$\gamma_1$	-1538.43	305.45	0.0002

$BA_1$  = Basal area  $ha^{-1}$ ;  $V_i$  = Volume  $ha^{-1}$ ;  $B_i$  = Biomass  $ha^{-1}$ ;  $X_1$  = Mid-infrared;  $\Psi$  = Value of the parameters;  $\epsilon^t$  = Standard error of the parameters,  $\infty$  = Significance level of the parameters.

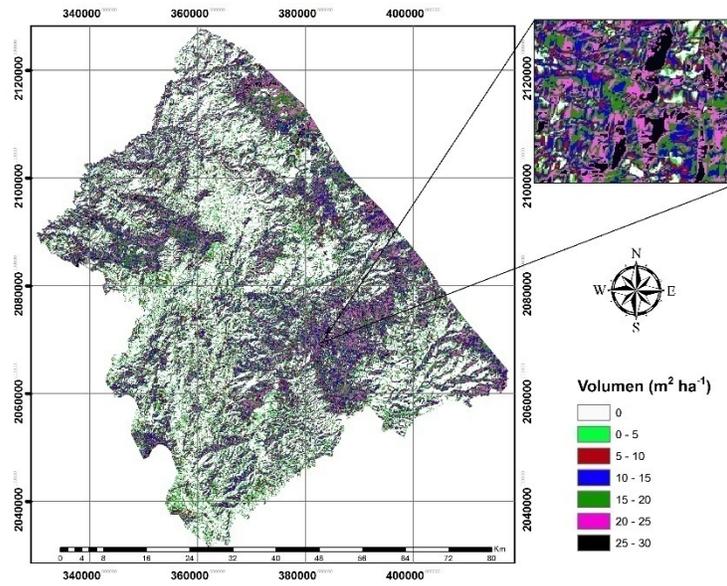
The five models presented good statistical bases to accurately estimate the forest variables (BA, V and B), which are based on the MIR band, the auxiliary variable (independent variables). However, the prominent models that best describe the biological behavior of the basal area, volume and biomass were models 1, 3 and 5, which obtained the desirable statistics, the highest value for the adjusted determination coefficient ( $R^2_{adj} = 0.52, 0.54$  y  $0.60$  for BA, V and B, respectively), and the lowest value for the root of the mean square error (RMSE =  $3.30, 13.18$  and  $5.91$  for BA, V and B, respectively). These results differ from those obtained by other authors; for example, Muñoz *et al.* (2014) register higher errors in BA and V, with

RMSE =  $6.70 \text{ m}^2 \text{ ha}^{-1}$  and  $41.45 \text{ m}^3 \text{ ha}^{-1}$  in the temperate forest, based on Spot 4 and 5 images. Aguirre *et al.* (2007) estimated the BA, V and B using Spot 5 images in *Pinus patula* Schiede ex Schltdl. & Cham. forests under management, and calculated an RSME of  $11.87 \text{ m}^2 \text{ ha}^{-1}$ ,  $96.81 \text{ m}^3 \text{ ha}^{-1}$  and  $52.56 \text{ Mg ha}^{-1}$ , for BA, V and B, respectively. For their part, Hall *et al.* (2006) estimated an RMSE of  $33.7 \text{ Mg ha}^{-1}$  and  $74.7 \text{ m}^3 \text{ ha}^{-1}$  for BA and V, with the support of Landsat ETM imaging. The results obtained by other authors differ from those of the present study, which may be due to the fact that tropical forests, particularly the low deciduous forest, have lower values for BA, V and B than temperate forests.

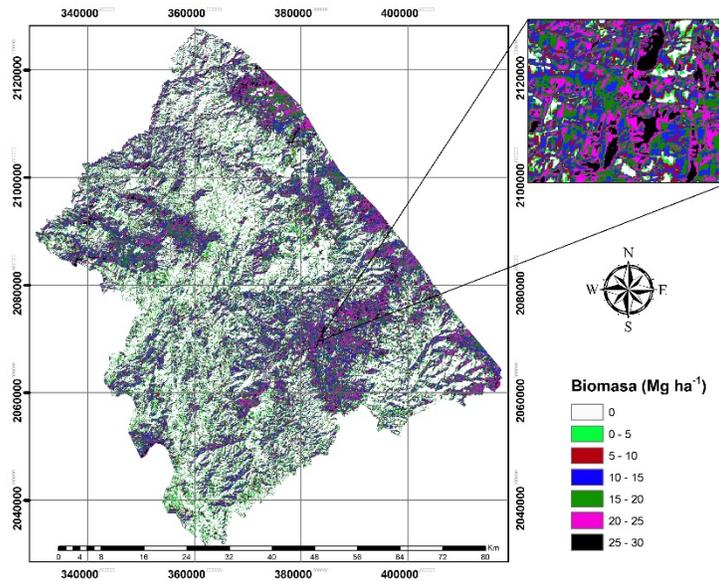
Figures 2, 3 and 4 correspond to the spatial distribution maps for the basal area ( $\text{m}^2 \text{ ha}^{-1}$ ), volume ( $\text{m}^3 \text{ ha}^{-1}$ ) and biomass ( $\text{Mg ha}^{-1}$ ) of the study area, respectively. The parts in white indicate the non-forest lands (areas without forest cover). Given the range of values for each variable, certain ranges pertain to a one-pixel surface area ( $900 \text{ m}^2$ ); therefore, their color is imperceptible, and only those areas where the values in which the variable is most abundant are apparent.



**Figure 2.** Spatial distribution of the basal area estimated with model 1.



**Figure 3.** Spatial distribution of the volume estimated with model 3.



**Figure 4.** Spatial distribution of the biomass estimated with model 5.



## Estimation of the forest parameters in terms of total inventories

Table 5 shows the estimations of the total inventory for basal area, volume and biomass obtained using the traditional method (in-field inventory) and the remote sensing method (regression estimators). The mid-infrared (MIR) band had the highest values for correlation with the forest parameters:  $BA = -0.72$ ,  $V_2 = -0.74$  and  $B = -0.77$ . Therefore, the regression estimators were built based on the MIR band for the calculation of the BA, V and B. The regression estimator showed the highest precision (permissible error under 10 %) compared to the simple random sampling. Statistically, the two methods produced similar results, although the regression estimator produces more conservative estimations and builds less ample confidence intervals in respect to the total inventory of the SRS (the most optimistic inventory).

**Table 5.** Comparison between the estimations of the variables ( $BA \text{ m}^2 \text{ ha}^{-1}$ ,  $V \text{ m}^2 \text{ ha}^{-1}$  and  $B \text{ Mg ha}^{-1}$ ) based on the traditional and the remote sensing (regression) inventories.

Inventory	Method	Estimators	BA $\text{m}^2$	V <sub>2</sub> $\text{m}^3$	B <sub>2</sub> Mg
Traditional	SRS	Inventory	6 387.54	23 665.11	11 567.92
		IC+95 %	6 453.52	23 935.51	11 697.57
		IC-95 %	6 321.56	23 394.70	11 438.28
Remote perception	RE	Inventory	6 149.63	22 668.06	11 066.06
		IC+95 %	6 196.81	22 856.46	11 150.58
		IC-95 %	6 102.45	22 479.66	10 981.53

$BA_1$  = Basal area;  $V_1$  = Volume;  $B_2$  = Biomass; SRS = Simple random sampling; RE = Regression estimator;  $CI_{\pm}$  = Confidence intervals at a level of reliability of 95 %.

The assessment of the forest resources with the support of remote sensing showed more conservative and more accurate results than the traditional inventory (SRS); this allows the forest managers to make better decisions for the sustainable management of the forests and enables them to be integrated in the future into pay-per-service environmental projects through the capture of carbon (Hall *et al.*, 2006; Aguirre *et al.*, 2007).

## Conclusions

Five regression models were generated –one for BA, two for V, and two for B–, and the mid-infrared band of the spectral data and vegetation indexes from Landsat 7 ETM+ images were used as auxiliary variables.

The combinations of data derived from a forest inventory (basal area, volume and biomass) and the satellite image variables (spectral bands and vegetation indices) through linear regression models produce spatial distribution maps for each of the forest parameters.

## Acknowledgements

The authors wish to express their gratitude to INIFAP for having authorized the Project "*Estimación de biomasa aérea en tres regiones forestales de México*" ("Estimation of the aboveground biomass in three forest regions of Mexico") and granted the corresponding Fiscal Funds in 2015 to finance its execution.

### **Conflict of interests**

The authors declare no conflict of interest.

### **Contribution by author**

Miguel Acosta Mireles: supervision of the data analysis and drafting of the original manuscript; Ramiro Pérez Miranda conceptualization and design of the study, drafting of the original manuscript; Martín Enrique Romero Sánchez: review of the data and analysis of the information, drafting of the original manuscript; Antonio González Hernández: revision and monitoring of the results, review and drafting of the manuscript; Luis Ángel Martínez: development of regression models and review and drafting of the final manuscript.

