

Estimating biomass of mixed and uneven-aged forests using spectral data and a hybrid model combining regression trees and linear models

Pablito M López-Serrano ⁽¹⁾, Carlos A López-Sánchez ⁽²⁾, Ramón A Díaz-Varela ⁽³⁾, José J Corral-Rivas ⁽²⁾, Raúl Solís-Moreno ⁽⁴⁾, Benedicto Vargas-Larreta ⁽⁵⁾, Juan G Álvarez-González ⁽⁶⁾

The Sierra Madre Occidental mountain range (Durango, Mexico) is of great ecological interest because of the high degree of environmental heterogeneity in the area. The objective of the present study was to estimate the biomass of mixed and uneven-aged forests in the Sierra Madre Occidental by using Landsat-5 TM spectral data and forest inventory data. We used the ATCOR3[®] atmospheric and topographic correction module to convert remotely sensed imagery digital signals to surface reflectance values. The usual approach of modeling stand variables by using multiple linear regression was compared with a hybrid model developed in two steps: in the first step a regression tree was used to obtain an initial classification of homogeneous biomass groups, and multiple linear regression models were then fitted to each node of the pruned regression tree. Cross-validation of the hybrid model explained 72.96% of the observed stand biomass variation, with a reduction in the RMSE of 25.47% with respect to the estimates yielded by the linear model fitted to the complete database. The most important variables for the binary classification process in the regression tree were the albedo, the corrected readings of the short-wave infrared band of the satellite (2.08-2.35 μm) and the topographic moisture index. We used the model output to construct a map for estimating biomass in the study area, which yielded values of between 51 and 235 Mg ha^{-1} . The use of regression trees in combination with stepwise regression of corrected satellite imagery proved a reliable method for estimating forest biomass.

Keywords: Regression Trees, Stepwise Regression, Remote Sensing, ATCOR3, Terrain Features, Image Texture

Introduction

The Sierra Madre Occidental is considered an area of special ecological interest because of the high levels of biodiversity, which are attributed to diverse physiographic and climatic conditions (Challenger 1998). The area is also important because of the presence of some of the most im-

portant commercial species of pine and oak in Mexican ecosystems (Sánchez et al. 2003).

Quantification of forest biomass and carbon sequestration is a relevant issue in the management of these forest stands. Reliable information is required for accurate biomass estimation, which should also take

into account variable external factors that can be modeled, e.g., climate change (IPCC 2003, Ryu et al. 2004, Sun & Ranson 2009). However, given the diversity of environmental, topographical and biophysical conditions in forest ecosystems in different locations, there is no universal, transferable technique for estimating biomass (Keller et al. 2001, Foody et al. 2003, Lu 2005, Cutler et al. 2012).

In general, forest biomass can be measured directly (destructive analysis) or it can be estimated indirectly (Brown & Lugo 1984). The direct method is usually accurate, but it is expensive and time-consuming and can only be used in small areas (Ketterings et al. 2001, Zianis & Mencuccini 2004, Walker et al. 2011). These difficulties have largely been resolved by the appearance and further development of quantitative satellite systems and aerial remote sensing, together with the development of parametric and nonparametric statistical methods for modeling variables of interest. The stand variables usually measured in traditional forest inventories can now be estimated faster, at lower cost and over larger areas (Liang 2007). The application of spatial technologies has allowed estimation of biomass in different ecosystems (Muukkonen & Heiskanen 2005, Fuchs et al. 2009, Hernández-Stefanoni et al. 2011, Aguirre-Salado et al. 2014).

□ (1) DICAF, Universidad Juárez del Estado de Durango, Boulevard del Guadiana 501, Ciudad Universitaria, Torre de Investigación, 34120 Durango, Dgo (México); (2) Instituto de Silvicultura e Industria de la Madera, Universidad Juárez del Estado de Durango, Boulevard del Guadiana 501, Ciudad Universitaria, Torre de Investigación, 34120 Durango, Dgo (México); (3) Departamento de Botánica - IBADER, Universidad de Santiago de Compostela, Escuela Politécnica Superior, Lugo (España); (4) Facultad de Ciencias Forestales, Universidad Juárez del Estado de Durango, Río Papaloapan 132, Valle del Sur Durango, 34120 Durango, Dgo (México); (5) División de Estudios de Posgrado e Investigación, Instituto Tecnológico de El Salto, Mesa del Tecnológico s/n, 34942, El Salto, Dgo (México); (6) Departamento de Ingeniería Agroforestal, Universidad de Santiago de Compostela, Escuela Politécnica Superior, Lugo (España)

@ Carlos A López-Sánchez (calopez@ujed.mx)

Received: Nov 17, 2014 - Accepted: May 17, 2015

Citation: López-Serrano PM, López-Sánchez CA, Díaz-Varela RA, Corral-Rivas JJ, Solís-Moreno R, Vargas-Larreta B, Álvarez-González JG (2015). Estimating biomass of mixed and uneven-aged forests using spectral data and a hybrid model combining regression trees and linear models. *iForest* 9: 226-234. - doi: 10.3832/ifor1504-008 [online 2015-09-21]

Communicated by: Davide Travaglini

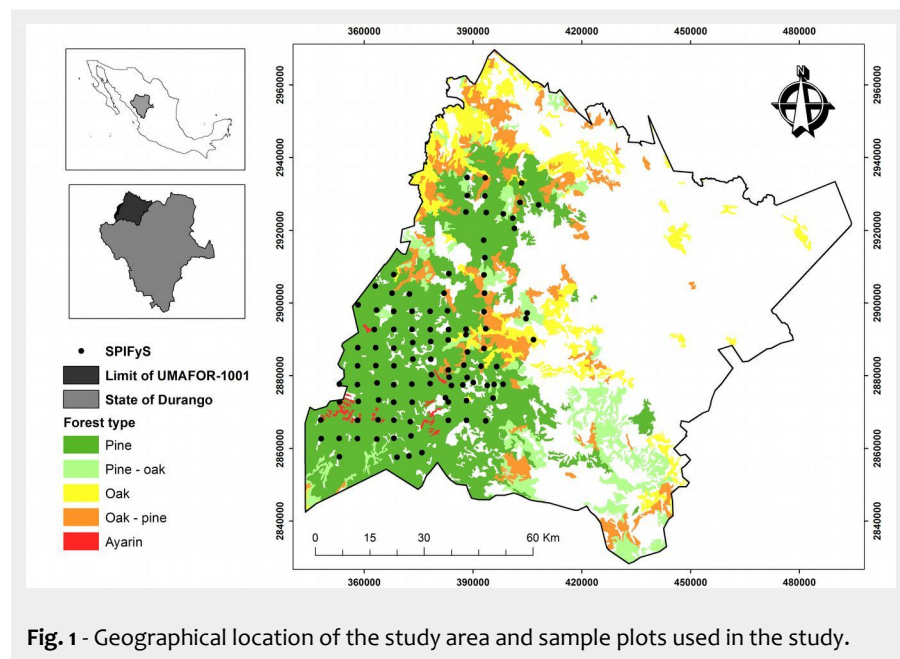


Fig. 1 - Geographical location of the study area and sample plots used in the study.

Pre-processing of satellite imagery is important for improving the quality and interpretation of data. Forest ecosystems generally cover rough terrain where the topographic conditions lead to variations in reflectance values because of the position of the sun (Meyer et al. 1993, Richter & Schlöpfer 2011, Richter 2013). Thus, the quality of the final product largely depends on accurate calibration of the sensors and on radiometric correction to minimize distortion and atmospheric effects (Li et al. 2010). In this respect, the use of atmospheric and topographic correction is therefore essential to counteract such effects, and digital elevation model (DEM) parameters such as slope, orientation, shadows cast, sky view and altitude can be used for such purposes (Balthazar et al. 2012, Richter 2013). These primary parameters, together with biophysical parameters such as vegetation indexes (Gilbert et al. 1997) and indexes derived by the analysis of the image texture (by quantification of the spatial variation in grey tones using a grey level co-occurrence matrix – GLCM), are very useful for identifying areas or objects of interest in the image (Haralick et al. 1973, Botero & Restrepo 2010). They can also be combined with terrain parameters to model vegetation characteristics (Lu 2005, Kayitakire et al. 2006, Díaz-Varela et al. 2011), as well as to describe hydrological, geomorphological and ecological processes

(Moore & Nieber 1989, Wilson & Galant 2000).

One of the methods most commonly used for this purpose is the classification and regression trees method, initially proposed by Breiman et al. (1984). This is a non-parametric, multivariate, supervised, inductive learning method that basically searches for classification and prediction rules by recursive partitioning. In this technique, a series of binary combinations (yes, no) expressed in terms of a single independent variable is used to identify certain profiles and vectors that enable description of the individual parameters under study (Hu et al. 2010).

The objective of the present study was to model the forest biomass in mixed and uneven-aged forests in the Sierra Madre Occidental (Mexico) by using remote sensing Landsat-5 TM imagery, terrain parameters and forest inventory data obtained from a network of permanent plots sampled in a traditional (ground based) survey. Two different approaches were compared: the usual modeling method of fitting a linear relationship to stand biomass and site variables obtained from remote sensing images, and a new approach consisting of a hybrid model combining regression trees and linear models for the final tree nodes. As far as we know, this is the first time this hybrid approach has been used to model stand biomass with remote sensing data.

Material and methods

Study area

The study area is located in the UMAFOR-1001 (*Unidad de Manejo Forestal Regional* or regional forest management unit) in the Sierra Madre Occidental, in the north of the state of Durango (Mexico), and covers an area of 1 142 916 ha (Fig. 1). The vegetation comprises pine, oak, Douglas fir, pine-oak and oak-pine forest, according to the description in the Land Use and Vegetation Cover Chart, scale 1:250.000, Series V (INEGI 2012).

Field data

A network of 99 permanent sampling plots was established during the winter of 2011, following the method described by Corral-Rivas et al. (2009). The plots were located by systematic sampling (with some exceptions to avoid non forested areas) with a grid of equidistant points separated by three to five kilometers, depending on the orography of the study area. In each plot (50 × 50 meters), the tree species were recorded and the diameter at breast height > 7.5 cm and total height (m) of all standing trees were measured.

Species-specific statistical models developed by Vargas-Larreta (2013) were used to estimate individual (at tree level) above-ground biomass. The goodness of fit for such statistical models ranged from 0.87 to 0.99 (R^2), and the root mean square error (RMSE) varied from 22.8 to 95.2 kg. Once the tree aboveground biomass was estimated, all values from each sampling plot were summed and expressed on a per hectare basis. Summary statistics including number of observations, mean, standard deviation, minimum, and maximum values of the main stand variables are shown in Tab. 1.

Datasets

Source of spectral data

Three Landsat-5 TM (Thematic Mapper) satellite images (path/row: 31/42, 32/41 and 32/42), obtained in April and May 2011 and covering the entire study area, were examined (available from the US Geological Service webpage - <http://glovis.usgs.gov/>). The available images are subjected to cubic convolution geometric correction for discrete data (level L1T), with a RMSE of the geometric residuals lower than 1 pixel, and they are therefore suitable for image processing (Keys 1981).

Landsat-5 TM data have spatial resolution of 30 m with a 16 day revisit period. The swath width is 185 km with seven spectral bands in the following wavelength regions of the electromagnetic spectrum: blue (0.45-0.52 μm), green (0.52-0.60 μm), red (0.63-0.69 μm), near infrared (0.78-0.89 μm), short-wave infrared (1.55-1.75 μm) and short-wave infrared (2.08-2.35 μm). These wavelength regions correspond respectively to bands 1, 2, 3, 4, 5 and 7 of the Landsat-

Tab. 1 - Statistics of the main stand variables. Dominant height was calculated as the mean height of the 100 thickest trees per hectare.

Variable	Mean	Standard deviation	Minimum value	Maximum value
Number of stems per ha	655.47	322.25	224.00	2264.00
Stand basal area ($\text{m}^2 \text{ha}^{-1}$)	20.30	6.42	8.22	41.12
Dominant height (m)	14.62	3.72	6.87	24.81
Stand biomass (Mg ha^{-1})	89.03	43.45	2.70	234.03

5 TM satellite (NASA 2011). Given its thermal characteristics, band 6 was not used in the present study.

Atmospheric and topographic correction (ATCOR3[®])

The satellite images were subjected to radiometrical, atmospheric and topographical correction using the ATCOR3[®] module, regarded as particularly suitable for mountainous zones (Geosystems 2013) and implemented in the ERDAS[®] IMAGINE[®] 2013 software (ERDAS 2013). After correction, the original image digital levels (DL) were converted to ground reflectance values for each band. A number of vegetation indexes and other derived parameters (Tab. 2) were computed from the atmospherically and topographically corrected image bands and then included in the biomass estimation models for their evaluation as possible regressor variables. Vegetation indexes are regarded as good indicators of vegetation cover “greenness” (understood as a combination of attributes such as leaf chlorophyll content, leaf area, canopy cover and structure – Glenn et al. 2008) and are good indicators of vegetated canopy biomass. Hence, the Normalized Difference Vegetation Index (Rouse et al. 1974) and Soil Adjusted Soil Vegetation Index (Huete 1988) were included as indexes correlated with green biomass content, with the former being particularly suited for scattered vegetation land cover. The Leaf Area Index (Baret & Guyot 1991) derived from NDVI was also included as a good indicator of green biomass. Albedo (Asrar 1989), photosynthetically active radiation (Asrar et al. 1984) and absorbed shortwave solar radiation (Brutsaert 1975) were also included as comprehensive indicators of the interaction between land cover and solar radiation in the visible and near-infrared regions of the electromagnetic spectrum.

The ATCOR3[®] module (Geosystems 2013) first calculates the radiance at sensor level ($Wsr^{-1} m^{-2}$) from the image pixels DL. Several input parameters were required for this calculation and were retrieved from the image metadata (*header file*): date of acquisition, scale factors, geometry (solar zenith angle and solar azimuth) and other information about the sensor calibration file (“gain and bias”). Other parameters were adjusted by taking into account particularities of the input datasets and the conditions of the imagery dates, e.g., visibility (35 km), pixel size of the DEM (15 m), aerosol type (rural), among others. As the image was cloudless and no suitable water vapor bands were available, dehazing/cloud removal and atmospheric water retrieval settings were retained as “default”, which, in this case, is recommended by the ATCOR3[®] User Manual (Geosystems 2013).

As a prior requisite for application of the ATCOR3[®] module, three topographic parameters (namely slope, orientation, skyview and shadows cast – Richter 2013) were computed from a DEM of the study area

Tab. 2 - Vegetation indexes analyzed in the present study. (NIR): Near-infrared band (0.83 μm); (RED): Red band (0.63 μm); (ρ): Reflectance; ($1-\rho(\lambda)$): Absorbed part of radiation; ($E_g(\lambda)$): the global (direct plus diffuse) solar flux on the ground; (C): Constant value 0.8; (A): Constant value 1; (B): Constant value 0.4; ($int_{0.3-2.5\mu m}$): extrapolation for region of the 0.3-2.5 μm (bands) of most satellite sensors; ($d\lambda$): adjustable parameter used to derive direct albedo on solar zenith angle.

Vegetation index	Definition	Author
Normalized Difference Vegetation Index	$NDVI = (NIR - RED) / (NIR + RED)$	Rouse et al. (1974)
Soil Adjusted Soil Vegetation Index	$SAVI = [(\rho_{NIR} - \rho_{RED}) \cdot 1.5] / [(\rho_{NIR} + \rho_{RED}) + 0.5]$	Huete (1988)
Modified Soil-adjusted Vegetation Index	$MSAVI2 = [2 \cdot NIR + 1 - \sqrt{(2 \cdot NIR + 1)^2 - 8 \cdot (NIR - RED)}] / 2$	Qi et al. (1994)
Leaf Area Index	$LAI = -(1/0.6) \ln[(0.6 - NDVI) / 0.78]$	Baret & Guyot (1991)
Albedo	$a = int_{0.3-2.5\mu m} [\rho(\lambda) d(\lambda)] / int_{0.3-2.5\mu m} (d\lambda)$	Asrar (1989)
Fraction of Photosynthetically Active Radiation	$FPAR = C - 1 - A \exp(-B \cdot LAI)$	Asrar et al. (1984)
Absorbed Shortwave Solar Radiation	$R_{solar} = int_{0.3-2.5\mu m} [1 - \rho(\lambda) E_g(\lambda) d\lambda]$	Brutsaert (1975)

with a spatial resolution of 15 m (INEGI 2014). Prior to these calculations, a low pass filter with a 5×5 moving window was applied to the original DEM in order to reduce the banding effects present in the original file.

After radiometric correction, the three scenes corresponding to the study area were mosaicked using the ERDAS[®] IMAGINE[®] 2013 software (ERDAS 2013).

Texture parameters

With the aim of including information that combines the spatial and spectral domain of the remote sensed imagery in the

biomass estimation models, the following texture parameters were calculated from the NDVI image based on grey level co-occurrence matrices (Tab. 3): homogeneity, contrast, dissimilarity, mean, standard deviation, entropy, second order angular moment and correlation (Haralick et al. 1973). Calculations were done at three different analysis scales, corresponding to window sizes of 3×3, 5×5 and 7×7 pixels, respectively. The original NDVI image values were resampled to a grey level depth of 256 (8 bits) to reduce computational costs (Haralick et al. 1973). This procedure was carried out using the software pack-

Tab. 3 - Additional variables for biomass modelling. (N): number of grey levels; (P): normalized symmetric GLCM of dimension N×N; (V): vector difference normalized grey level of dimension N; P(i, j): matrix of co-occurrence normalized, so that $\sum_{(i,j=0;N-1)} P(i,j) = 1$; V(k): normalized grey level difference vector $\sum_{(i,j=0;N-1)} \sum_{|i-j|=k} P(i,j)$; (Z): Average elevation; (R): Point radio altitude units; (As): Drainage area specified; ($\tan(\beta)$): Local slope angle; (VA): Variance; (ME): Mean. D, F, G and H were derived according to equation of Zevenbergen & Thorne (1987).

Group variable	Variable	Formula	Reference
Texture (NDVI)	Homogeneity (HO)	$HO = \sum_{(i,j=0;N-1)} i [P_{i,j} / 1 + (i-j)^2]$	Haralick et al. (1973)
	Contrast (CO)	$CO = \sum_{(i,j=0;N-1)} i P_{i,j} (i-j)^2$	
	Dissimilarity (DI)	$DI = \sum_{(i,j=0;N-1)} i P_{i,j} (i-j)$	
	Mean (ME)	$ME = \sum_{(i,j=0;N-1)} i P_{i,j}$	
	Standard Deviation (Std)	$Std = \sqrt{VA}$	
	Entropy (EN)	$EN = \sum_{(i,j=0;N-1)} i P_{i,j} [-\ln(i) - P_{i,j}]$	
	Angular Second Moment (ASM)	$ASM = \sum_{(i,j=0;N-1)} i P_{i,j}^2$	
Terrain (DEM)	Correlation (CR)	$CR = \sum_{(i,j=0;N-1)} i P_{i,j} [(i-ME)(j-ME) / \sqrt{(VA_i VA_j)}]$	-
	Elevation	Digital Elevation Model	
	Slope (β)	$\beta = \arctan(G^2 + H^2)^{1/2}$	
	Transformed Aspect (Trasp)	$Trasp = 1 - \cos[(\pi / 180)(\alpha - 30)] / 2$	
	Terrain Shape Index (TSI)	$TSI = Z / R$	
	Wetness Index (WI)	$W = \ln[As / \tan(\beta)]$	
	Profile curvature (\emptyset)	$\emptyset = -2[DG^2 + EH^2 + FGH] / (G^2 + H^2)$	
Plan curvature (ω)	$\omega = 2[DH^2 + EG^2 + FGH] / (G^2 + H^2)$		
Curvature (x)	$x = \omega - \emptyset$	Wilson & Gallant (2000)	

age PCI Geomatica 2013[®] (PCI Geomatics 2013).

Terrain variables

Terrain variables are directly related to forest species composition, tree height growth, and other forest stand variables, and enable these to be modeled (McNab 1989, Roberts & Cooper 1989). Therefore, first and second order terrain parameters (Tab. 3) were derived from the 5×5 low pass filtered DEM (INEGI 2014) and included as candidate variables in the models. The selected first order terrain parameters were elevation, slope, transformed aspect, profile curvature, plan curvature and curvature, while second order terrain parameters were terrain shape index and wetness index. These parameters are potentially related to key features for forest stand development, such as overall climate characteristics, insolation, evapotranspiration, run-off, infiltration, wind exposure and site productivity. Some of these terrain features are widely used in hydrological, geomorphological and ecological studies (Wilson & Gallant 2000), whereas others are used more specifically for vegetation and forest assessment (McNab 1989, Roberts & Cooper 1989).

Dataset integration

The sample plots were geopositioned with the aim of extracting the pixel value average with an associated buffer of 25 m for each potential predictor. This extraction was carried out using the statistical software R (R Core Team 2014) and the “raster” package. Finally, a database was constructed with the mean biomass values for each plot: the corrected bands of the Landsat-5 TM sensor (6 bands: 1, 2, 3, 4, 5 and 7), the vegetation indexes (6 indexes), the texture variables derived from the Normalized Difference Vegetation Index (NDVI – 24 variables), and the terrain variables derived from the DEM (8 variables).

Models fitted

The biomass of the sample plots was estimated using two different methods. In the first, the ordinary least squares (OLS) method was used to fit a linear model to estimate stand biomass. The best set of independent variables was selected by using the stepwise variables selection method. The second method consisted of a two-step hybrid approach. In the first step, a regression tree was used to classify the sample plots in homogeneous groups according to their biomass values by a binary rule-based method. In the second step, the ordinary least square method was used to fit linear models to estimate stand biomass for each group by using the stepwise variable selection method to select the best set of independent variables. In both methods, the 45 spectral, texture and terrain variables were taken into account as possible independent variables.

Regression tree analysis (CaRT) is a non-

parametric technique for the sequential partitioning of a data set composed of a continuous response variable and any number of potential continuous or categorical predictor variables, by using dichotomous criteria (Breiman et al. 1984). After each split, the algorithm identifies the predictor variable that provides the most effective binary separation of the range in the response variable. As a result, predictor variables can be used more than once. The regression tree analysis was performed using the “rpart” package in R (Therneau & Atkinson 2012, R Core Team 2014). This approach partitions the data set sequentially, considering two-way splits at each tree node. The best split at each node t is the split that maximizes the following quantity (eqn.1):

$$\Delta Err(s,t) = Err(t) - P_L Err(t_L) - P_R Err(t_R)$$

where P_L and P_R are the proportions of sample plots that fall respectively to the left and right branch of node t ; $Err(t_L)$ and $Err(t_R)$ are the error of the left and right branches; $Err(t)$ is the mean square error at node t given by (eqn. 2):

$$\frac{1}{N_t} \cdot \sum_{i=1}^{N_t} (y_i - \bar{y}_t)^2$$

and \bar{y}_t is the stand biomass assigned to node t , calculated as the mean of the stand biomass of all the sample plots in node t .

Instead of applying stopping rules, a sequence of sub-trees was generated by growing a large tree and pruning it back until only the root node was left. The error of each sub-tree was then estimated by cross-validation, and the sub-tree with the lowest error was chosen by analyzing the values of the complexity parameter defined by Breiman et al. (1984).

Once the sample plots of each final node were obtained, a multiple linear model was fitted to estimate the stand biomass, using stepwise selection methods to select the best set of independent variables, with the SAS/STAT[®] software package (SAS Institute Inc 2007). Two criteria were considered for the evaluation of model performances: the coefficient of determination (R^2) and the RMSE. The expressions of these statistics are summarized as follows (eqn. 3, eqn. 4):

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

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

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

The main problem associated with such multiple linear models is the multicollinea-

arity. This refers to the existence of high correlations between certain independent variables representing or measuring similar phenomena. Although the least-squares estimates of regression coefficients remain unbiased and consistent under the presence of multicollinearity, they are no longer efficient (Myers 1990). This may seriously affect the standard errors of the coefficients, thus invalidating statistical tests and confidence intervals (Neter et al. 1990). One of the main sources of multicollinearity is the use of overfitted models that include several polynomial and cross-product terms. To evaluate the presence of multicollinearity between variables in the models, the condition number, defined as the square root of the ratio of the largest to the smallest eigenvalue of the correlation matrix, was used. According to Belsey (1991), condition numbers between 5 and 10 indicate that multicollinearity will not be a major problem, while those in the range 30-100 indicate moderate multicollinearity and those in the range 1000-3000 indicate severe multicollinearity. Therefore, independent variables with condition numbers higher than 30 were not used in the models.

Finally, since the quality of fit does not necessarily reflect its predictive performance (Myers 1990), an assessment of the validity of the models with an independent dataset is recommended (Kozak & Kozak 2003). Due to the difficulties associated with collecting such data, cross-validation was applied in this study. Validation of the model fitted to the complete database (method 1) and of the model of each final node was thus based on the values of coefficient of determination and root mean square error, using the predicted residual sum of squares (PRESS), *i.e.*, each sample plot is removed in turn and the model is refitted using the remaining sample plots. The out-of-sample predicted value is calculated for the omitted sample plot in each case, and the PRESS statistic is calculated as the sum of the squares of all the resulting prediction errors.

The equations obtained with the best method were finally used to generate a map of biomass by means of the map algebra and conditional tools of the GIS package ArcGIS 10[®] (ESRI 2012) from the vector vegetation layer (INEGI 2012).

Results

The parameter estimates of the linear model fitted to the complete database using OLS and the stepwise variables selection method is shown in Tab. 4. All the parameters were significant at a 5% level, and up to 5 independent variables were included in the model without generating multicollinearity problems. The model explained 58.83% of the observed stand biomass variability with a RMSE of 27.88 Mg ha⁻¹ (31.32% of the mean stand biomass). Based on the results of cross-validation, the model explained 51.33% of the total

observed variation in stand biomass with a RMSE value of 30.31 Mg ha^{-1} (34.04% of the mean stand biomass).

The regression tree shown in Fig. 2 was generated in the first step of the second methods. This tree has a root node that contains all 99 sample plots with an assigned mean biomass value of 89.03 Mg ha^{-1} . The limiting value of 121.5, for the variable albedo, divided these samples into two groups of plots. Each subgroup was then sequentially divided by the limiting values of the variables band 7, band 5, LAI, contrast texture with a 5×5 window and wetness index.

However, the problem with the regression trees method is that it tends to overfit the data, and therefore the most general model may not be obtained when a new set of independent data is used (Breiman et al. 1984). These authors suggested that once the tree is constructed, it should be exhaustively pruned by successively removal of branches or terminal nodes that contribute little to explaining the response variable, to yield an appropriately-sized

Tab. 4 - Model obtained for the total database by OLS with stepwise selection of independent variables (*): RMSE expressed as a percentage of mean stand biomass.

Parameter	Estimate	Standard error	OLS		Cross-validation	
			RMSE (Mg ha^{-1})	R^2	RMSE (Mg ha^{-1})	R^2
Intercept	572.4939	172.2304				
SAVI	-0.3673	0.1566				
Band 7	-2.7166	0.6893				
Abs. Shortw. solar rad.	0.0797	0.0245	27.88 (31.32%)*	0.5833	30.31 (34.04%)*	0.5133
LAI	0.0166	0.0037				
Modified SAVI	-2215.8205	717.3984				
NDVI	2072.8881	628.6811				
Wetness index	3.4027	1.5666				
Contrast 7×7	0.2073	0.0926				

tree. The mean value of the complexity parameter (CP) defined by Breiman et al. (1984) and obtained by cross-validation, was used in this case to select the number of branches on the final tree, and the result is shown in Fig. 3 (the number of tree input variables was reduced to three, namely albedo, band 7 and wetness index).

Direct application of the regression tree

to the data from the 99 permanent sample plots used in this study resulted in 56.76% of the observed variability in stand biomass being explained by the model, with an RMSE value of 28.57 Mg ha^{-1} (32.09% of the mean stand biomass). Once the four groups shown in Fig. 3 were obtained from the tree, linear models were fitted to each. The parameter estimates, their standard

Fig. 2 - Classification tree obtained by the regression tree method. n is the number of sample plots in each node and W is the biomass value for each node (Mg ha^{-1}).

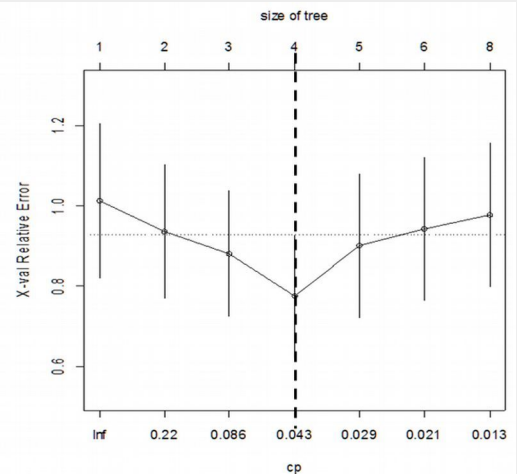
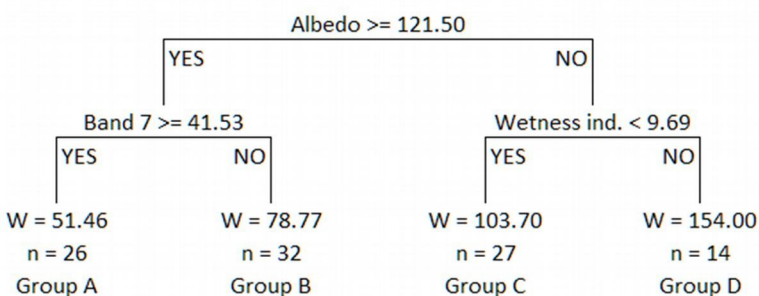
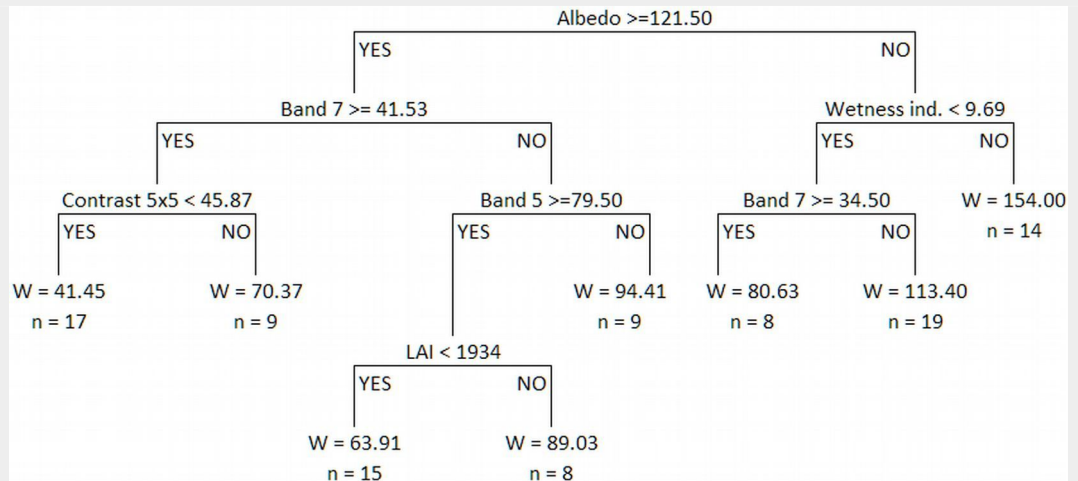


Fig. 3 - Classification tree obtained by pruning the regression tree (left) and plot of the relationships between the cost-complexity parameter (CP), the cross validation error (x-val Relative Error) and tree size (number of nodes). The dashed vertical line represents the maximum number of nodes (corresponding cost-complexity parameter) for which the cross validation error is greater than the standard error (right).

Tab. 5 - Model obtained for the final nodes of the regression tree by OLS and stepwise selection of independent variables (*): RMSE expressed as a percentage of mean stand biomass; (**): percentage difference between the cross-validation RMSE of the hybrid model compared with the same statistic obtained by cross-validation of the linear model fitted to the complete database.

Group	Parameter	Estimate	Standard error	Cross-validation		RMSE reduction** (%)
				RMSE (Mg ha ⁻¹)	R ²	
A	Abs. Shortw. solar rad.	0.1004	0.0523	19.85 (38.57%)*	0.2605	14.73%
	Entropy 3×3	-143.8093	32.8994			
	Correlation 3×3	-35.0613	11.0553			
	Dissimilarity 5×5	12.3776	2.9020			
	Mean 7×7	1.6886	0.4856			
B	Band 1	4.5965	0.7375	16.87 (21.42%)*	0.2591	18.97%
	Band 5	-2.1912	0.5013			
C	Intercept	360.9605	102.8189	26.71 (25.76%)*	0.1382	1.05%
	Band 1	-4.7552	1.9528			
	Correlation 3×3	-137.3135	41.1643			
D	Flow solar rad.	-0.2223	0.0667	31.38 (20.38%)*	0.6149	46.14%
	NDVI	745.2892	112.9641			
	Stand. Dev. 7×7	-7.0717	2.0254			

errors and the goodness-of-fit statistics obtained by cross-validation are shown in Tab. 5.

The intercepts of models for groups A, B and D were not significant at the 5% level, and therefore the models were refitted without this term. In all cases, the parameters were significant and the condition number values indicate no problems associated with multicollinearity. Analysis of the graph of the residuals plotted against the predicted values also indicated the absence of problems associated with variance heterogeneity or lack of normal distribution of the residuals (Fig. 4).

Cross-validation of the hybrid model comprising the pruned regression tree (Fig. 3) and the linear models for each terminal node explained 72.96% of the observed variation in stand biomass, with a RMSE value of 22.59 Mg ha⁻¹ (25.37% of the mean stand biomass).

The spatial distribution of the biomass estimation (Mg ha⁻¹) in the study area obtained by application of the classification rules included in the regression tree model and the posterior estimations yielded by multiple linear regression are shown in Fig. 5. The grey and green pixels reflect biomass contents lower than 80 Mg ha⁻¹ and between 80 and 157 Mg ha⁻¹, respectively, whereas the yellow and red pixels represent the highest biomass values found in temperate forest (mainly pine and pine-oak cover) in the study area, in accordance with the INEGI's Land Use and Vegetation chart, series V (INEGI 2012).

Discussion

The results of the present research showed that integration of spectral information, texture variables derived from the NDVI and terrain indexes (DEM) was essential for forest biomass estimation. Indeed, these variables were reported in previous studies as being closely related to the development and growth of this type of ecosystem and are also useful for ecosystem evaluation and monitoring (McNab 1993, Chen et al. 2004, Díaz-Varela et al. 2011).

The combined use of regression trees and linear models including spectral, texture and terrain variables proved to be a good method for identifying patterns and defining biomass trends in the study area. The independent variable albedo, defined as the average solar reflectance (Liang 2000), was the main discriminating factor in the regression tree, and highest values occurred in the areas with the lowest forest biomass. Kuusinen et al. (2014) obtained similar results and observed an inverse relationship between stand age and albedo, so that the value of this variable was lower in mature stands because of the lower level of incident radiation absorbed in such stands. This relationship can be used to discriminate zones with different levels of forest biomass. The other two discriminant variables were spectral band 7 (short-wave infrared) and the topographical wetness in-

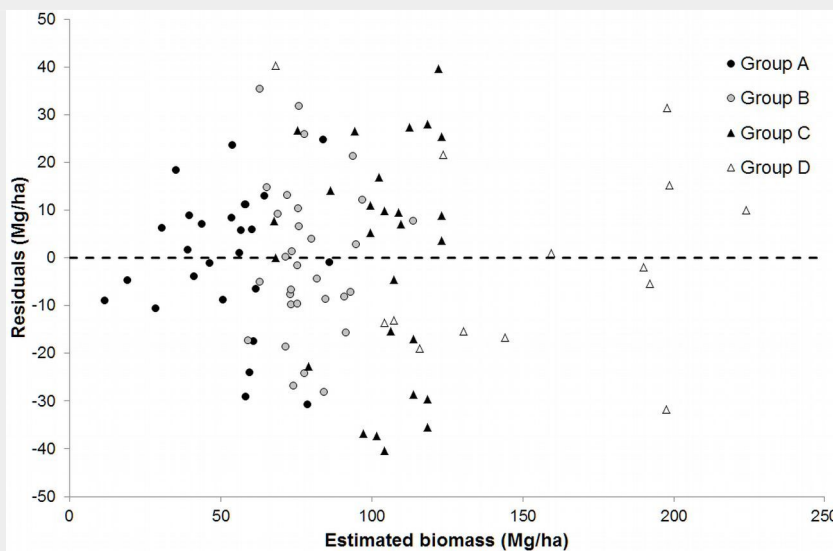


Fig. 4 - Plot of residual values against estimated biomass for groups obtained from the classification tree.

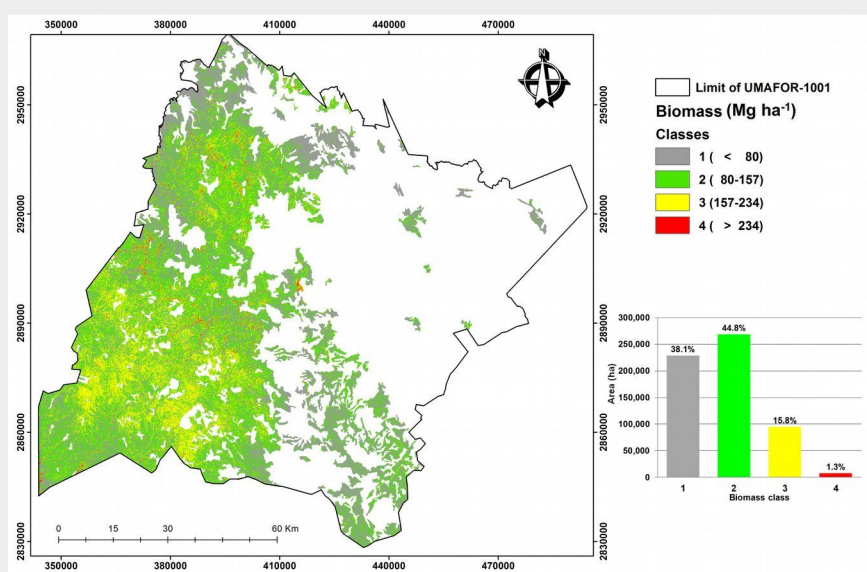


Fig. 5 - Biomass distribution in the study area.

dex. Because of its spectral characteristics, band 7 is directly related to the moisture content of soil and vegetation recorded by the image. Thus, the reflectivity in this band increased as the surface wetness captured by the sensor decreased. In afforested areas, this band displays low reflectivity; as moisture levels are high in forest stands, the highest values for this band represent lower amounts of biomass in the classification tree. These results are similar to those reported by García et al. (2005) for pure stands of *Pinus halepensis* and *P. sylvestris* in Spain, i.e., there was an inverse relationship between the values for band 7 of the Landsat-5 TM sensor and the moisture content of the residual forest biomass. Finally, inclusion of the topographic wetness index in the model confirms the previous findings, as high values of this index are associated with high levels of moisture, which coincide with zones with high amounts of biomass. In various studies, use of the relationship between the wetness index and the vegetation biomass has enabled identification of the distribution of vegetation (Moore et al. 1993, Zinko et al. 2005) and of potential areas for establishing forest plantations (Holmgren 1994). These results indicate water availability as a key factor controlling biomass production in arid and semi-arid environments such as the Sierra Madre Occidental (Salinas-Zavala et al. 2002, Méndez-Barroso et al. 2009, Zhao & Running 2010, Forzieri et al. 2014).

The hybrid model combining the nonparametric method of regression trees and multiple linear models yielded a reduction in the RMSE (25.47%) and an increase in R^2 (42.14%) with respect to the same statistics obtained by cross-validation of the linear model fitted to the complete database. According to the results shown in Tab. 5, the main RMSE reduction was obtained in group D (46.14%), probably because this group is associated with sample plots with higher stand biomass values. On the other hand, the reduction in RMSE in group C was only 1.05%, possibly because this is the group with the lowest coefficient of variation of stand biomass (27.23% compared with a mean value of 48.56%).

The parameters selected by the hybrid model included single band values in the visible (Band 1-blue) and mid infrared (Band 5) regions of the electromagnetic spectrum. The mid infrared regions have already been included and discussed in the initial (Band 5 and Band 7) and pruned (Band 7) regression tree model, indicating a relationship between the biomass and spectral response of forest cover in the imagery bands related to water content. As expected, vegetation indexes, such as NDVI, and other indicators of the radiation-land cover interaction, such as the Absorbed Shortwave Solar Radiation, also emerged as valuable predictors of biomass due to their potential relationship with biomass. The relationship between remote

sensed NDVI and biomass content, which has been the matter of discussion as strongly dependent on the scale of analyses and characteristics of the imagery, has nonetheless been regarded in the literature as one of the most widely used predictors of biomass content (Foody et al. 2003, González-Alonso et al. 2006).

Interestingly, apart from these variables in the pure spectral domain, up to five variables of the spectral-spatial domain (i.e., texture variables Entropy 3×3, Correlation 3×3, Dissimilarity 5×5, Mean 7×7, Stand. Dev. 7×7) were included in the mixed model. This indicates the importance of the spatial arrangement of spectral values at different spatial scales (from a 3×3 kernel corresponding with an area of 0.81 ha to a 7×7 kernel corresponding with an area of 4.41 ha) for forest stand characterization, as reported in previous studies (Franklin et al. 2001, Kayitakire et al. 2006, Díaz-Varela et al. 2011, Nichol & Sarker 2011).

The value of R^2 finally obtained with the hybrid model (0.7296) is slightly higher than that obtained by Sun et al. (2011) in a study carried out in the US, with high resolution LIDAR sensors and SAR data, to model field-measured biomass by linear models and stepwise selection of variables (R^2 , 0.71 and RMSE, 31.33 Mg ha⁻¹).

Estimates obtained with sensors of medium spatial resolution usually display a low predictive power for each band of the sensor. Thus, Foody et al. (2003) found the strongest predictive relationship for biomass with a sampling network specifically designed for different sites ($r > 0.71$) based on indexes obtained for tropical forest in Brazil by using Landsat TM data. On the other hand, Houghton et al. (2007) estimated the biomass of Russian forests by using data derived from a MODIS sensor and regression trees in 500×500 m plots, in which the percentage of variance explained by regression trees ranged from 1 to 67%.

In the present study, the consideration of biophysical variables derived from satellite images along with other complementary data and the use of nonparametric multivariate techniques, improved the quality of the estimates, thus indicating that this is a promising line of research.

Conclusions

This study explored possible improvements in forest biomass prediction involving use of field data and geodata derived from atmospherically and topographically corrected satellite images (provided by the Landsat-5 TM sensor), texture indexes and DEM-derived terrain variables. A new approach combining the nonparametric regression trees method and multiple regression analysis of the groups defined in the pruned tree was compared with the usual method of fitting a linear model to the complete database. Cross-validation of both methods indicated that the proposed new approach improved the performance

of the linear model. Moisture content was an important covariate in the final model and was directly related to biomass distribution in the temperate forest under study. The proposed approach deserves further attention in future studies aimed at estimating stand variables by using remote sensing data, especially for more complicated stand structures, such as mixed and uneven aged forests, in which the use of a mean value for each node cannot accurately represent the intra-node stand variation.

Acknowledgements

This research was supported by SEP-PROMEP (Project: *Seguimiento y Evaluación de Sitios Permanentes de Investigación Forestal y el Impacto Socioeconómico del Manejo Forestal en el Norte de México*). We thank three anonymous reviewers for their useful comments and suggestions, which helped to improve an earlier version of the manuscript.

References

- Aguirre-Salado CA, Aguirre-Salado EJ, Treviño-Garza OA, Aguirre-Calderón J, Jiménez-Pérez MA, González-Tagle JR, Valdéz-Lazalde G, Sánchez-Díaz RH, Aguirre-Salado AI, Miranda-Aragón L (2014). Mapping aboveground biomass by integrating geospatial and forest inventory data through a k-nearest neighbour strategy in North Central Mexico. *Journal of Arid Land* 1: 80-96. - doi: [10.1007/s40333-013-0191-x](https://doi.org/10.1007/s40333-013-0191-x).
- Asrar G (1989). *Theory and applications of optical remote sensing*. John Wiley and Sons, New York, USA, pp. 734. [online] URL: <http://www.cabdirect.org/abstracts/19910745919.html>
- Asrar G, Fuchs M, Kanemasu ET, Hatfield JL (1984). Estimating absorbed photosynthetically active radiation and leaf area index from spectral reflectance in wheat. *Agronomy Journal* 76 (2): 300-306. - doi: [10.2134/agronj1984.00021962007600020029x](https://doi.org/10.2134/agronj1984.00021962007600020029x)
- Balthazar V, Veerle V, Eric FL (2012). Evaluation and parameterization of ATCOR3 topographic correction method for forest cover mapping in mountain areas. *International Journal of Applied Earth Observation and Geoinformation* 18: 436-450. - doi: [10.1016/j.jag.2012.03.010](https://doi.org/10.1016/j.jag.2012.03.010)
- Baret F, Guyot G (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment* 35: 161-173. - doi: [10.1016/0034-4257\(91\)90009-U](https://doi.org/10.1016/0034-4257(91)90009-U)
- Belsey DA (1991). *Conditioning diagnostics, collinearity and weak data in regression*. John Wiley and Sons Inc, New York, USA, pp. 396.
- Botero VJS, Restrepo MA (2010). Análisis de textura en panes usando la matriz de coocurrencia [Analysis of texture in breads using the co-occurrence matrix]. *Revista Politécnica* 6 (10): 74-80. [in Spanish]
- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984). *Classification and regression trees*. Chapman and Hall, New York, USA, pp. 368.
- Brown S, Lugo AE (1984). Biomass of tropical forests: a new estimate based on forest volumes. *Science* 223 (4642): 1290-1293. - doi: [10.1126/science.223.4642.1290](https://doi.org/10.1126/science.223.4642.1290)
- Brutsaert W (1975). On a derivable formula for

- long-wave radiation from clear skies. *Water Resources Research* 11: 742-744. - doi: [10.1029/WR011005p00742](https://doi.org/10.1029/WR011005p00742)
- Challenger A (1998). Utilización y conservación de los ecosistemas terrestres de México: pasado, presente y futuro [Use and conservation of terrestrial ecosystems of Mexico: past, present and future]. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad/Instituto de Biología, UNAM/Agrupación Sierra Madre, México, DF, pp. 847. [in Spanish]
- Chen D, Stow DA, Gong P (2004). Examining the effect of spatial resolution and texture window size on classification accuracy: an urban environment case. *International Journal of Remote Sensing* 25 (11): 2177-2192. - doi: [10.1080/01431160310001618464](https://doi.org/10.1080/01431160310001618464)
- Corral-Rivas J, Vargas B, Wehenkel C, Aguirre O, Alvarez JG, Rojo A (2009). Guía para el establecimiento de sitios de inventario periódico forestal y de suelos del estado de Durango [Guidelines for the establishment of permanent sample plots in forests of Durango State]. Facultad de Ciencias Forestales, Universidad Juárez del Estado de Durango, Mexico, pp. 81. [in Spanish]
- Cutler MEJ, Boyd DS, Foody GM, Vetrivela A (2012). Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: an assessment of predictions between regions. *ISPRS Journal of Photogrammetry and Remote Sensing* 70: 66-67. - doi: [10.1016/j.isprsjprs.2012.03.011](https://doi.org/10.1016/j.isprsjprs.2012.03.011)
- Díaz-Varela RA, Álvarez P, Díaz-Varela E, Calvo S (2011). Prediction of stand quality characteristics in sweet chestnut forests in NW Spain by combining terrain attributes spectral textural features and landscape metrics. *Forest Ecology and Management* 261: 1962-1972. - doi: [10.1016/j.foreco.2011.02.023](https://doi.org/10.1016/j.foreco.2011.02.023)
- ERDAS (2013). *Erdas Imagine 2013*. Hexago AB, Intergraph Corporation, Madison, AL, USA. [online] URL: <http://www.hexagongeospatial.com/products/ERDAS-IMAGINE/details.aspx>
- ESRI (2012). *ArcGIS for Desktop 10*. Redwoods, CA, USA. [online] URL: <http://www.esri.com/software/arcgis/arcgis-for-desktop>
- Foody GM, Boyd DS, Cutler MEJ (2003). Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment* 85 (4): 463-474. - doi: [10.1016/S0034-4257\(03\)00039-7](https://doi.org/10.1016/S0034-4257(03)00039-7)
- Forzieri G, Feyen L, Cescatti A, Vivoni ER (2014). Spatial and temporal variations in ecosystem response to monsoon precipitation variability in southwestern North America. *Journal of Geophysical Research: Biogeosciences* 119 (10): 1999-2017. - doi: [10.1002/2014JG002710](https://doi.org/10.1002/2014JG002710)
- Franklin SE, Maudie AJ, Lavigne MB (2001). Using spatial co-occurrence texture to increase forest structure and species composition classification accuracy. *Photogrammetric Engineering and Remote Sensing* 67: 849-856. [online] URL: http://asprs.org/a/publications/pers/2001journal/july/2001_jul_849-855.pdf
- Fuchs H, Magdon P, Kleinn C, Flessa H (2009). Estimating aboveground carbon in a catchment of the Siberian forest tundra: combining satellite imagery and field inventory. *Remote Sensing of Environment* 113: 518-531. - doi: [10.1016/j.rse.2008.07.017](https://doi.org/10.1016/j.rse.2008.07.017)
- García MA, Pérez CF, De la Riva J, Fernández J, Pascual PE, Herranz A (2005). Estimación de la biomasa residual forestal mediante técnicas de teledetección y SIG en masas puras de *Pinus halepensis* y *P. sylvestris* [Estimation of residual forest biomass using remote sensing technology and GIS in pure stands of *Pinus halepensis* and *P. sylvestris*]. In: Proceedings of the "IV Congreso Forestal Español de la Sociedad Española de Ciencias Forestales". Paper no. 4CFE05-342-T1, 308, Sociedad Española de Ciencias Forestales, Palencia, Spain, pp. 5. [in Spanish]
- Geosystems (2013). Haze reduction, atmospheric and topographic correction. User Manual ATCOR2 and ATCOR3, Geosystems GmbH, Gernering, Germany, pp. 238. [online] URL: <http://www.geosystems.de/atcor/>
- Gilbert MA, Piqueros JG, Haro JG (1997). Acerca de los índices de vegetación [About vegetation indices]. *Revista de Teledetección* 8: 1-10. [in Spanish]
- Glenn EP, Huete AR, Nagler PL, Nelson SG (2008). Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: what vegetation indices can and cannot tell us about the landscape. *Sensors* 8: 2136-2160. - doi: [10.3390/s8042136](https://doi.org/10.3390/s8042136)
- González-Alonso F, Merino-De-Miguel S, Roldán-Zamarrón A, García-Gigorro S, Cuevas JM (2006). Forest biomass estimation through NDVI composites. The role of remotely sensed data to assess Spanish forests as carbon sinks. *International Journal of Remote Sensing* 27: 5409-5415. - doi: [10.1080/01431160600830748](https://doi.org/10.1080/01431160600830748)
- Haralick RM, Shanmugam K, Dinstein I (1973). Textural features for image classification. *IEEE Transactions of Systems, Man, and Cybernetics* 6: 610-621. - doi: [10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314)
- Hernández-Stefanoni JL, Gallardo-Cruz JA, Meave JA, Dupuy JM (2011). Combining geostatistical models and remotely sensed data to improve tropical plant richness mapping. *Ecological Indicators* 11: 1046-1056. - doi: [10.1016/j.ecolind.2010.11.003](https://doi.org/10.1016/j.ecolind.2010.11.003)
- Holmgren P (1994). Topographic and geochemical influence on the forest site quality, with respect to *Pinus sylvestris* and *Picea abies* in Sweden. *Scandinavian Journal of Forest Research* 9: 75-82. - doi: [10.1080/02827589409382815](https://doi.org/10.1080/02827589409382815)
- Houghton RA, Butman D, Bunn AG, Krankina ON, Schlesinger P, Stone TA (2007). Mapping Russian forest biomass with data from satellites and forest inventories. *Environmental Research Letters* 2 (4): 045032. - doi: [10.1088/1748-9326/2/4/045032](https://doi.org/10.1088/1748-9326/2/4/045032)
- Hu W, Mengersen K, Tong S (2010). Risk factor analysis and spatiotemporal CART model of cryptosporidiosis in Queensland, Australia. *BMC Infectious Diseases* 10 (1): 311. [online] URL: <http://www.biomedcentral.com/1471-2334/10/311/>
- Huete AR (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25: 295-309. - doi: [10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- INEGI (2012). *Uso del suelo y vegetación escala 1:250.000* [Land use and vegetation scale 1:250.000]. Serie V, Información vectorial, Instituto Nacional de Estadística Geográfica e Informática, México. [in Spanish] [online] URL: <http://www.inegi.org.mx/geo/contenidos/recnat/usuarios/>
- INEGI (2014). Continuo de elevaciones mexicano 3.0 (CEM 3.0) [Mexican continuous elevation 3.0 (CEM3.0)]. Instituto Nacional de Estadística Geográfica e Informática, México. [in Spanish] [online] URL: <http://www.inegi.org.mx/geo/contenidos/datosrelieve/continental/descarga.aspx>
- IPCC (2003). Good practice guidance for land use, land-use change and forestry. IPCC National Greenhouse Gas Inventories Programme, Hayama, Japan, pp. 590.
- Kayitakire F, Hamel C, Defourny P (2006). Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery. *Remote Sensing of Environment* 102 (3-4): 390-401. - doi: [10.1016/j.rse.2006.02.022](https://doi.org/10.1016/j.rse.2006.02.022)
- Keller M, Palace M, Hurtt G (2001). Biomass estimation in the Tapajos National Forest, Brazil: examination of sampling and allometric uncertainties. *Forest Ecology and Management* 154 (3): 371-382. - doi: [10.1016/S0378-1127\(01\)00509-6](https://doi.org/10.1016/S0378-1127(01)00509-6)
- Ketterings QM, Coe R, van Noordwijk M, Ambagau Y, Palm CA (2001). Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests. *Forest Ecology and Management* 146 (1-3): 199-209. - doi: [10.1016/S0378-1127\(00\)00460-6](https://doi.org/10.1016/S0378-1127(00)00460-6)
- Keys RG (1981). Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 29: 1153-1160. - doi: [10.1109/TASSP.1981.1163711](https://doi.org/10.1109/TASSP.1981.1163711)
- Kozak A, Kozak RA (2003). Does cross validation provide additional information in the evaluation of regression models? *Canadian Journal of Forest Research* 33: 976-987. - doi: [10.1139/x03-022](https://doi.org/10.1139/x03-022)
- Kuusinen N, Tomppo E, Shuai Y, Berninger F (2014). Effects of forest age on albedo in boreal forests estimated from MODIS and Landsat albedo retrievals. *Remote Sensing of Environment* 145: 145-153. - doi: [10.1016/j.rse.2014.02.005](https://doi.org/10.1016/j.rse.2014.02.005)
- Li F, Jupp D.L, Reddy S, Lymburner L, Mueller N, Tan P, Islam, A (2010). An evaluation of the use of atmospheric and BRDF correction to standardize Landsat data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3 (3): 257-270. - doi: [10.1109/JSTARS.2010.2042281](https://doi.org/10.1109/JSTARS.2010.2042281)
- Liang S (2000). Narrowband to broadband conversions of land surface albedo I algorithms. *Remote Sensing of Environment* 76: 213-238. - doi: [10.1016/S0034-4257\(00\)00205-4](https://doi.org/10.1016/S0034-4257(00)00205-4)
- Liang S (2007). Recent developments in estimating land surface biogeophysical variables from optical remote sensing. *Progress in Physical Geography* 31: 501-516. - doi: [10.1177/0309133307084626](https://doi.org/10.1177/0309133307084626)
- Lu D (2005). Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon Basin. *International Journal of Remote Sensing* 26 (12): 2509-2525. - doi: [10.1080/01431160500142145](https://doi.org/10.1080/01431160500142145)
- McNab WH (1989). Terrain shape index: quantifying effect of minor landforms on tree height. *Forest Science* 35: 91-104. [online] URL: <http://www.ingentaconnect.com/content/saf/fs/1989/00000035/00000001/art00010>

- McNab HW (1993). A topographic index to quantify the effect of mesoscale landform on site productivity. *Canadian Journal of Forest Research* 23: 1100-1107. - doi: [10.1139/x93-140](https://doi.org/10.1139/x93-140)
- Méndez-Barroso LA, Vivoni ER, Watts CJ, Rodríguez JC (2009). Seasonal and interannual relation between precipitation, surface soil moisture and vegetation dynamics in the North American monsoon region. *Journal of Hydrology* 377 (1-2): 59-70. - doi: [10.1016/j.jhydrol.2009.08.009](https://doi.org/10.1016/j.jhydrol.2009.08.009)
- Meyer P, Itten KI, Kellenberger T, Sandmeier S, Sandmeier R (1993). Radiometric corrections of topographically induced effects on Landsat TM data in an Alpine environment. *ISPRS Journal of Photogrammetry and Remote Sensing* 48 (4): 17-28. - doi: [10.1016/0924-2716\(93\)90028-L](https://doi.org/10.1016/0924-2716(93)90028-L)
- Moore ID, Nieber JL (1989). Landscape assessment of soil erosion and non-point source pollution. *Journal of the Minnesota Academy of Science* 55 (1): 18-24.
- Moore ID, Norton TW, Williams JE (1993). Modeling environmental heterogeneity in forested landscapes. *Journal of Hydrology* 150 (2-4): 717-747. - doi: [10.1016/0022-1694\(93\)90133-T](https://doi.org/10.1016/0022-1694(93)90133-T)
- Muukkonen P, Heiskanen J (2005). Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. *Remote Sensing of Environment* 99: 434-447. - doi: [10.1016/j.rse.2005.09.011](https://doi.org/10.1016/j.rse.2005.09.011)
- Myers RH (1990). *Classical and modern regression with applications* (2nd edn). Duxbury Press, Belmont, CA, USA, pp. 488.
- NASA (2011). *Landsat 7 science data users handbook*. National Aeronautics and Space Administration, NASA, Washington, DC, USA, pp. 30-31. [online] URL: http://landsathandbook.gsfc.nasa.gov/pdfs/Landsat7_Handbook.pdf
- Neter J, Wasserman W, Kutner MH (1990). *Applied linear statistical models: regression, analysis of variance and experimental designs* (3rd edn). Irwin, Boston, MA, USA, pp. 842.
- Nichol JE, Sarker MLR (2011). Improved biomass estimation using the texture parameters of two high-resolution optical sensors. *IEEE Transactions on Geoscience and Remote Sensing* 49: 930-948. - doi: [10.1109/TGRS.2010.2068574](https://doi.org/10.1109/TGRS.2010.2068574)
- PCI Geomatics (2013). *PCI Geomatics 2013*. PCI Geomatics Inc, Ontario, Canada. [online] URL: <http://www.pcigeomatics.com/>
- Qi J, Chehbouni A, Huete AR, Kerr YH (1994). Modified soil adjusted vegetation index (MSAVI). *Remote Sensing of Environment* 48 (2): 119-126. - doi: [10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1)
- R Core Team (2014). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria, pp. 3551. [online] URL: <http://www.R-project.org/>
- Richter R (2013). *User manual ATCOR2 and ATCOR3, ATCOR for IMAGINE 2013: haze reduction, atmospheric and topographic correction*. DLR Oberpfaffenhofen, Institute of Ptoelectronics, D-82234, Version 15.01.2013, Wessling, Germany, pp. 240.
- Richter R, Schläpfer D (2011). Atmospheric/topographic correction for satellite imagery. *ATCOR 2/3 User Guide, Version 8.0.2, ReSe, Wil, Switzerland*, pp. 252. [online] URL: <http://www.researchgate.net/publication/268376667>
- Roberts DW, Cooper SV (1989). Concepts and techniques of vegetation mapping. In: "Land classifications based on vegetation: applications for resource management" (Ferguson D, Morgan P, Johnson FD eds). Gen. Tech. Rep. INT-257, Intermountain Forest and Range Experiment Station, USDA Forest Service, Ogden, UT, USA, pp. 90-96.
- Rouse JW, Haas RH, Schiell JA, Deferino DW, Harlan JC (1974). *Monitoring the vernal advancement of retrogradation of natural vegetation*. NASA/OSFC, Type III Final Report, Remote Sensing Center, Texas A&M University, College Station, TX, USA, pp. 87. [online] URL: <http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19750020419.pdf>
- Ryu S, Chen J, Crow TR, Saunders SC (2004). Available fuel dynamics in nine contrasting forest ecosystems in North America. *Environmental Management* 33 (1): 87-107. - doi: [10.1007/s00267-003-9120-7](https://doi.org/10.1007/s00267-003-9120-7)
- Salinas-Zavala CA, Douglas AV, Diaz HF (2002). Interannual variability of NDVI in northwest Mexico: associated climatic mechanisms and ecological implications. *Remote Sensing of Environment* 82 (2-3): 417-430. - doi: [10.1016/S0034-4257\(02\)00057-3](https://doi.org/10.1016/S0034-4257(02)00057-3)
- Sánchez O, Vega E, Peters E, Monrroy-Vilchis O (2003). *Conservación de ecosistemas templados de montaña en México [Conservation of temperate mountain ecosystems in Mexico]*. Instituto Nacional de Ecología, INE-SEMARNAT, México, DF, pp. 315. [in Spanish]
- SAS Institute Inc (2007). *SAS OnlineDoc (version 9.2)*. SAS Institute Inc, Cary NC, USA. [online] URL: http://www.sas.com/en_us/software/sas9.html
- Sun G, Ranson JK (2009). Radiometric slope correction for forest biomass estimation from SAR data in the Western Sayani Mountains, Siberia. *Remote Sensing of Environment* 79 (2-3): 279-287.
- Sun G, Ranson KJ, Guo C, Zhang Z, Montesano P, Kimes D (2011). Forest biomass mapping from lidar and radar synergies. *Remote Sensing of Environment* 115: 2906-2916 - doi: [10.1016/j.rse.2011.03.021](https://doi.org/10.1016/j.rse.2011.03.021)
- Therneau TM, Atkinson B (2012). Package "rpart" (version 3:1-52). Web site. [online] URL: <http://cran.r-project.org/web/packages/rpart/index.html>
- Vargas-Larreta B (2013). *Estimación del potencial de los bosques de Durango para la mitigación del cambio climático. Modelización de la biomasa forestal [Assessing the potential of forests of Durango for mitigating climate change. Modelling of forest biomass]*. Proyecto FOMIX-DGO-2011-C01-165681, Comisión Nacional de Ciencia y Tecnología, CONACYT, Mexico, pp. 53. [in Spanish]
- Walker W, Baccini A, Nepstad M, Horning N, Knight D, Braun E, Bausch A (2011). *Guía de campo para la estimación de biomasa y carbono forestal [Field guide to estimate forest biomass and carbon]* (version 1.0). Woods Hole Research Center, Falmouth, MA, USA, pp. 53. [in Spanish]
- Wilson JP, Gallant JC (2000). Digital terrain analysis. In: "Terrain Analysis: Principles and Applications" (Wilson JP, Gallant JC eds). John Wiley and Sons, Inc, New York, USA, pp. 1-28. [online] URL: <http://cnrfiles.uwsp.edu/turyk/Database/Development/MJB/>
- Zevenbergen LW, Thorne CR (1987). Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms* 12: 47-56. - doi: [10.1002/esp.3290120107](https://doi.org/10.1002/esp.3290120107)
- Zhao M, Running SW (2010). Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* 329 (5994): 940-943. - doi: [10.1126/science.1192666](https://doi.org/10.1126/science.1192666)
- Zianis D, Mencuccini M (2004). On simplifying allometric analyses of forest biomass. *Forest Ecology and Management* 187(2): 311-332. - doi: [10.1016/j.foreco.2003.07.007](https://doi.org/10.1016/j.foreco.2003.07.007)
- Zinko U, Seibert J, Dynesius, M, Nilsson C (2005). Plant species numbers predicted by a topography based groundwaterflow index. *Ecosystems* 8: 430-441. - doi: [10.1007/s10021-003-0125-0](https://doi.org/10.1007/s10021-003-0125-0)