

Modelling dasometric attributes of mixed and uneven-aged forests using Landsat-8 OLI spectral data in the Sierra Madre Occidental, Mexico

Carlos A López-Sánchez⁽¹⁾,
Pedro García-Ramírez⁽²⁾,
Richard Resl⁽³⁾,
José C Hernández-Díaz⁽¹⁾,
Pablito M López-Serrano⁽²⁾,
Christian Wehenkel⁽¹⁾

Remote sensors can be used as a robust and effective means of monitoring isolated or inaccessible forest sites. In the present study, the multivariate adaptive regression splines (MARS) technique was successfully applied to remotely sensed data collected by the Landsat-8 satellite to estimate mean diameter at breast height ($R^2 = 0.73$), mean crown cover ($R^2 = 0.55$), mean volume ($R^2 = 0.57$) and total volume per plot ($R^2 = 0.41$) in the forest monitoring sites. However, the spectral data yielded poor estimates of tree number per plot ($R^2 = 0.22$), the mean height ($R^2 = 0.25$) and the mean diameter at base ($R^2 = 0.38$). Seven spectral bands (band 1 to band 7), six vegetation indexes and other derived parameters (NDVI, SAVI, LAI, FPAR, ALB and ASR) and eight terrain variables derived from the digital elevation model (elevation, slope, aspect, plan curvature, profile curvature, transformed aspect, terrain shape index and wetness index) were used as predictors in the fitted models. To prevent over-parameterization only some of the predictor variables considered were included in each model. The results indicate the MARS technique is potentially suitable for estimating dasometric variables from using spectral data obtained by the Landsat-8 OLI sensor.

Keywords: Multivariate Adaptive Regression Splines, Mixed Forest, Uneven-aged Forest, Stand Variables, Remote Sensing, Terrain Features

Introduction

The Sierra Madre Occidental (SMO) mountain range is of great ecological interest due to its environmental heterogeneity, which is attributed to the broad physiological and climatic diversity in the area. Moreover, the SMO shelters uneven-aged and species-rich pine oak forests that are economically important ecosystems for Mexico and other parts of the world. 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 covers a total area of 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 this land (0.5 million), affected by warmer climate, is occupied by forest classified as rainforest (SRNyMA-CONAFOR 2007). The state of Durango generates between 25% and 30% of the national timber production, producing a total of 1.5 million m³ of roundwood per year, and promotes forest reserves as important sources of environmental services (López-Serrano et al. 2016).

Estimation of diverse dasometric vari-

ables in forest stands is a fundamental aspect of forest inventories. Commercial volume is particularly important in forest management and is valuable for the conservation of diverse forest ecosystems and for improving the economic productivity of forests (Gadow et al. 2004). Other stand variables are also important in forest management, such as diameter at breast height, total height, crown cover, basal area and plant biomass. Stand variables are usually measured on the ground by forest inventory teams in field surveys covering large areas (Trotter et al. 1997); however, such surveys are both expensive and time-consuming (Gleason & Im 2011).

Use of remote sensors is a cost-effective method of collecting data in large inaccessible areas and provides an alternative source of data for extrapolating and estimating forest variables (Franklin et al. 2003, Poulain et al. 2010). The data obtained can be rapidly updated and compared with existing data, and the method can be easily integrated with Geographical Information Systems. These methods have been successfully combined in several studies, and statistically significant correlations between the remotely-sensed data (captured by Landsat satellite sensors) and field-measured data have been reported for diameter at breast height, age and total height (Donoghue et al. 2004), volume (Hall et al. 2006, Poulain et al. 2010) and biomass (Luther et al. 2006, Ji et al. 2012, Zhang et al. 2014). The aim of the present study was to analyze the relationship be-

□ (1) Instituto de Silvicultura e Industria de la Madera. Universidad Juárez del Estado de Durango, Blvd del Guadiana 501, Fracc. Ciudad Universitaria, C.P. 34120 Durango (México); (2) Doctorado Institucional en Ciencias Agropecuarias y Forestales. Universidad Juárez de Estado de Durango (UJED); (3) GEOcentro UNIGIS-USFQ. Universidad de San Francisco de Quito, Cumbaya (Ecuador)

@ Christian Wehenkel (wehenkel@ujed.mx)

Received: Oct 01, 2015 - Accepted: Oct 12, 2016

Citation: López-Sánchez CA, García-Ramírez P, Resl R, Hernández-Díaz JC, López-Serrano PM, Wehenkel C (2017). Modelling dasometric attributes of mixed and uneven-aged forests using Landsat-8 OLI spectral data in the Sierra Madre Occidental, Mexico. *iForest* 10: 288-295. - doi: 10.3832/ifor1891-009 [online 2017-02-11]

Communicated by: Alessandro Montaghi

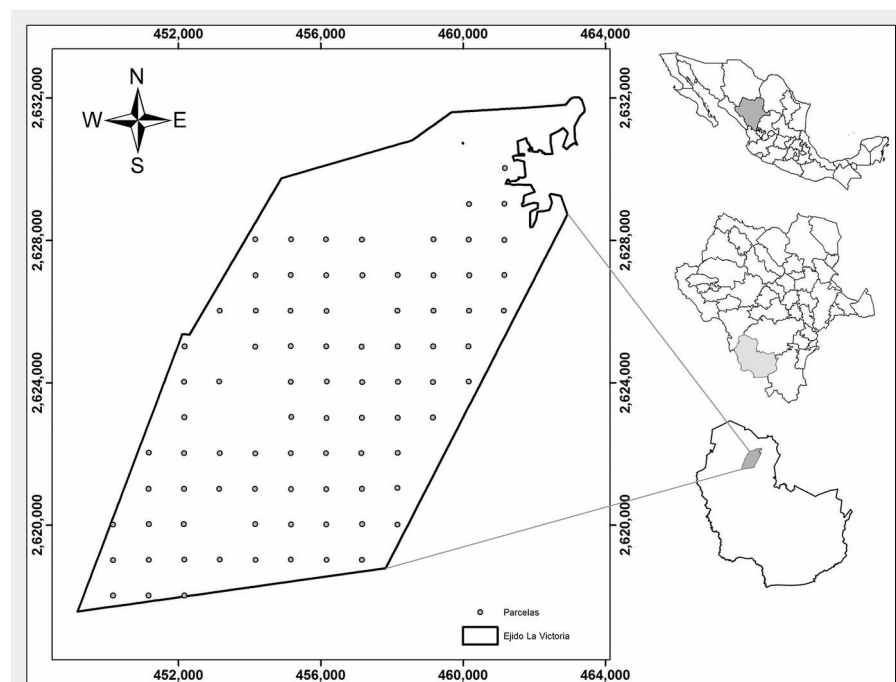


Fig. 1 - Maps showing the location of the study area and sample plots used in the study: (right, top to bottom) the State of Durango in Mexico; the municipality of Pueblo Nuevo in the State of Durango; the Ejido La Victoria in the municipality of Pueblo Nuevo. The area covered by the Ejido La Victoria is outlined in the large inset.

tween a set of field-measured dasometric variables and spectral data obtained by the Landsat-8 OLI (Operational Land Imager) sensor and data derived from a digital elevation model, aimed at estimating the total number of trees per plot, crown cover, mean diameter at base (0.30 m), mean diameter at breast height (1.30 m), mean height, mean volume and total volume per plot in a mixed uneven-aged forests in the Sierra Madre Occidental.

Material and methods

Study area

The study was carried out in the Ejido La Victoria (surface area: 10,810 ha), located in the Sierra Madre Occidental in the municipality of Pueblo Nuevo, state of Durango, Mexico (Fig. 1). Site elevation ranges between 2245 and 2870 m above sea level, according to the data obtained from the digital elevation model (DEM) developed

by the National Institute of Geography and Statistics of Mexico (INEGI 2014). The forests in the study area show a rich biodiversity and include at least 12 coniferous tree species (of which 10 are *Pinus* species) and 5 species of *Quercus*. The dominant types of vegetation in the area are mixed conifer forest and pine-oak forest (Wehenkel et al. 2011, González-Elizondo et al. 2012). The main species present are: *Pinus cooperi*, *P. durangensis*, *P. engelmannii*, *P. leiophylla*, *P. strobiformis*, *P. teocote*, *P. herrerae*, *Juniperus deppeana*, *Quercus sideroxil* and *Q. crasifolia* (Martínez-Antúnez et al. 2015).

Field sampling

A systematic sampling method (1 × 1 km between plots) was used to locate the sampling plots in the study area. In total, 83 circular sampling plots each of surface area 1000 m² were established. The plots were georeferenced relative to a reference system (WGS 84 Datum / UTM zone 13N). The plot data were obtained with the aid of a submeter GNSS receiver, tree calipers (Haglof, Sweden), hypsometer (Vertex IV[®], Haglof, Sweden) and metric tape. The following variables were measured in each plot, taking into account all the trees with diameter at breast height (DBH) larger than 5 cm: tree number (N), diameter at base (DB in cm, measured at 0.30 m above the ground), DBH (in cm, measured at 1.30 m above the ground), total height (HT in m) and crown diameter (CD in m). The volume (V in m³) per plot was estimated using specific allometric equations developed by Corral-Rivas et al. (2007) for the same study area; the descriptive statistics corresponding to the response variables are shown in Tab. 1.

Datasets

Source of spectral data

The image used in the study was captured by the Landsat-8 OLI sensor (LC80310442014018LGN00) and supplied by the USGS Global Visualization Viewer Server (<http://glovis.usgs.gov/>). The image was captured at about the same time as the field sampling was carried out (January, 2014). The images were obtained as a L1T level product (i.e., they were geometrically corrected with ground control points and the DEM). The satellite image was digitally preprocessed by radiometric correction and atmospheric and topographic techniques, following the procedures established in the ATCOR3[®] module (Atcor for Erdas Imagine – ERDAS 2013) developed for mountainous areas and implemented with ERDAS[®] IMAGINE[®] 2013 software (ERDAS 2013). The digital levels (DL) of each of the bands (Tab. 2) were converted to surface reflectance values by reading the model parameters and the orbit of the geometrical satellites included in the image metadata. Bands 1 to 7 (B1, B2, B3, B4, B5, B6 and B7) of Landsat-8 OLI were used in the present study.

Tab. 1 - Descriptive statistics for the response variables in each of 83 plots (of surface area, 1000 m²). (STD): standard deviation.

Stand variable	Mean	STD	Min	Max
Tree number per plot (N)	79.53	58.02	7.00	430.00
Mean diameter at base 0.30 m (cm - DB mean)	21.40	6.41	10.79	41.24
Mean diameter at breast height 1.30 m (cm - DBH mean)	17.96	5.79	8.05	35.45
Mean height (m - H mean)	10.95	3.17	6.01	23.80
Mean crown diameter (m - CD mean)	3.85	0.91	1.88	6.54
Mean volume (m ³ - V mean)	0.42	0.30	0.06	1.68
Total volume (m ³) per plot (TV)	24.98	11.78	1.31	67.14

Tab. 2 - Spectral bands, wavelength, spatial resolution and abbreviated name of the Landsat-8 OLI sensor.

Band number	Wavelength (μm)	Spatial resolution	Abbreviation
Band 1 - Coastal aerosol	0.433-0.453	30 m	B1
Band 2 - Blue	0.450-0.515	30 m	B2
Band 3 - Green	0.525-0.600	30 m	B3
Band 4 - Red	0.630-0.680	30 m	B4
Band 5 - Near Infrared (NIR)	0.845-0.885	30 m	B5
Band 6 - Short-Wave Infrared (SWIR) 1	1.560-1.660	30 m	B6
Band 7 - Short-Wave Infrared (SWIR) 2	2.100-2.300	30 m	B7

A number of vegetation indexes and other derived parameters (normalized difference vegetation index, *NDVI*; soil adjusted vegetation index, *SAVI*; leaf area index, *LAI*; fraction of photosynthetically active radiation, *FPAR*; albedo, *ALB*; absorbed shortwave solar radiation, *ASR*) were computed from the atmospherically and topographically corrected image bands and then included in the stand variables estimation models for evaluation as possible regressor variables (Tab. S1 in Supplementary material).

Terrain variables

These variables have influence on forest species composition, tree height growth, and other forest stand variables (McNab 1989, Roberts & Cooper 1989). First order variables included: elevation (*DEM*), slope (*SLOPE*), transformed aspect (*TRASP*), profile curvature (*PROFCURV*), plan curvature (*PLANCURV*) and curvature (*CURV*); and second order variables were: terrain shape index (*TERRSHPIN2*) and wetness index (*WETIND*). Terrain parameters were generated in the software ArcGIS 10[®] (ESRI 2012), with a 5×5 low pass filtered DEM (INEGI 2014) and included as candidate variables in the models (Tab. S2 in Supplementary material). These parameters are potentially related to key features affecting 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 18-meter radius for geolocalization of the plots. The values were extracted using the raster package available in the R statistical software (R Core Team 2015). Finally, a database was constructed with the mean stand variables values for each plot: the corrected bands of the Landsat-8 OLI sensor (7 bands), the vegetation indexes (6 indexes) and the terrain variables derived from the DEM (8 variables).

Statistical analysis

As the Shapiro-Wilks test revealed that the data were not normally distributed, the multivariate adaptive regression splines (MARS) technique was used for model construction. The MARS technique is an innovative, flexible method that constructs models for continuous, categorical or binary-type dependent and/or independent variables. It is a non-parametric regression technique based on constructing flexible models by fitting data to linear regressions in segments, combination of which gener-

ates local models where the relationship between the predictive and response variables may be linear or non-linear and which also incorporate interactions between variables. Each segment of the regression line represents a stretch of linear response to the variable and is denominated the basic function; the final model consists of the entire set of basic functions. The model can be represented so that it separately identifies additive contributions and those associated with different multivariate interactions (Friedman 1991). The general form of the MARS non parametric regression model, formulated on the dependent y variable and the x predictors, is as follows (eqn. 1):

$$y = f(x) + \varepsilon$$

where ε is the error, and $f(x)$ is the unknown regression function, derived as follows (eqn. 2):

$$f(x) = \beta_0 + \sum_{m=1}^M \beta_m B_m(x)$$

where β_0 is the intercept of the model, $B_m(x)$ is the m^{th} basis function, β_m is the coefficient of the m^{th} basis function, and M is the number of basis functions in the model. Each basis function $B_m(x)$ takes one of the following two forms: (i) a hinge function of the form $\max(0, x-k)$ or $\max(0, k-x)$, where k is a constant value; (ii) a product of two or more hinge functions, which can therefore model the interaction between two or more predictors (x).

In this study, MARS analysis was performed using the “earth” package in R (R Core Team 2015). To improve predictions and prevent over-fitting, MARS uses a modified form of the “cross-validation method” (*GCV* – Vidoli 2011). *GCV* is an approximation of the error that would be determined by leave-one-out validation and considers both the residual error and the model complexity. Therefore, the optimal model is that yielding the lowest *GCV* (eqn. 3):

$$GCV(M) = \frac{\sum_{i=1}^n [y_i - f_m(x_i)]^2}{n(1 - p_M/n)^2}$$

where y_i represents the observed values of the dependent variable, n is the number of observations, $f_m(x_i)$ is the MARS model with M basis functions, x_i represents the observed values of the predictors included in the MARS model and p_M is the number of parameters of the MARS model.

The importance of the measurement variable was also considered, to take into account whether model information was used or not. Use of the previously mentioned *GCV* parameter, together with the parameters N of the subsets (N_{sub}) and the residual sum of squares (*RSS*), yields reliable results. The N_{sub} criterion considers the number of model subsets that include the variable, i.e., one subset for each model

size from one up to the selected model size (in the final pruned model). The wrapper model utilizes the performance of the algorithm MARS to determine which features should be selected. The subset that produces the highest overall accuracy is deemed the best. For the *RSS* criterion, the decrease in the *RSS* is first calculated for each subset relative to the previous subset. For each variable, the decreases are then summed over all subsets that include that variable (Alonso-Fernández et al. 2014).

Three criteria were used to evaluate the model performance: (i) the determination coefficient (R^2) as a measure of how well the model fits the training data; (ii) the generalized determination coefficient (GR^2), which is based on the *GCV* statistic and the root mean square error (*RMSE*) for each model to provide a generalization of the predictive power of the model; and (iii) the common method of k -fold cross-validation (for $k = 2$ to 10 folds), in which error statistics are calculated across all k trials. Additional information on MARS can be found in Friedman (1991) and Hastie et al. (2009).

Cross validation was used with the aim of defining a dataset to test the model in the training phase, to minimize problems such as overfitting and to predict how the model will perform with an independent dataset. This provides a good indication of how well the classifier will perform with unseen data and also indicates the degree of variability in the k -fold cross-validation results.

For the purpose of comparison, multiple linear regression (MLR) was also carried out to compare how it performs relative to the MARS technique for modeling stand variables in the study area. MLR is the technique most commonly used in this kind of studies, as it is easy to understand and widely used in most scientific disciplines (Fassnacht et al. 2014). Stepwise selection (forward and backward) was performed to select the most informative variables that were included in the model using the stepAIC function and the exact Akaike’s information criterion (AIC) in the MASS package (Venables & Ripley 2002). The aim of this procedure was to identify the model with the smallest AIC by removing or adding predictors.

The equations obtained with the MARS technique (R^2 and *RMSE*) were finally used to generate a map of the stand variables by means of the map algebra and conditional tools of the GIS package ArcGIS 10[®] (ESRI 2012).

Results

Tab. 3 shows the results produced by the MLR and MARS techniques for each of the dasometric variables considered in the study. Although MARS generally performed best, MLR yielded better prediction of the *H mean* variable.

Tab. 4 shows the results obtained with the MARS models that performed best for estimating the dasometric variables consid-

Tab. 3 - Comparison of the best MLR and MARS model predicting seven tree stand attributes based on the coefficient of determination (R^2), root mean square error (RMSE) and RMSE percent. (N): the number of trees per plot; (DB mean): mean diameter at 0.30 m (cm); (DBH mean): mean diameter at 1.30 m (cm); (H mean): mean height (m); (CD mean): mean crown diameter (m); (V mean): mean volume (m^3); (TV): total volume (m^3) per plot; (MLR): MLR technique; (MARS): MARS technique.

Variable	MLR				MARS			
	Predictors	R^2	RMSE	%RMSE	Predictors	R^2	RMSE	%RMSE
N	5	0.18	54.20	68.15	4	0.22	51.01	64.14
DB mean	6	0.33	5.47	25.56	3	0.38	5.02	23.48
DBH mean	6	0.31	4.99	27.79	10	0.73	3.01	16.79
H mean	7	0.36	2.65	24.20	1	0.25	2.73	24.96
CD mean	6	0.32	0.78	20.31	8	0.55	0.61	15.83
V mean	7	0.30	0.27	64.09	7	0.57	0.20	47.39
TV	5	0.40	9.39	37.59	3	0.41	9.00	36.02

Tab. 4 - MARS model selection criteria for dasometric variables considered. (N): number of trees per plot; (DB mean): mean diameter at 0.30 m (cm); (DBH mean): mean diameter at 1.30 m (cm); (H mean): mean height (m); (CD mean): mean crown diameter (m); (V mean): mean volume (m^3); (TV): total volume (m^3) per plot; (GCV): generalized cross-validation; (RSS): residual sum of squares; (GR^2): generalized coefficient of determination.

Variable	Number of terms	GCV	RSS	GR^2
N	5 of 22	3273.51	215972.50	0.04
BD mean	6 of 34	33.55	2095.71	0.19
DBH mean	17 of 34	26.66	754.26	0.26
H mean	2 of 36	8.04	619.94	0.21
CD mean	9 of 21	0.59	30.85	0.30
V mean	9 of 34	0.06	3.24	0.34
TV	4 of 22	96.60	6722.15	0.31

ered in each plot of mixed and uneven aged forest plot in the study area.

According to the selection criteria, the model used to estimate the mean diameter at breast height (1.30 m), mean volume and the mean crown diameter were those that performed best with the training data set (i.e., highest $R^2 = 0.73, 0.57$ and 0.55 ; $GR^2 = 0.26, 0.34$ and 0.30 ; and lowest RMSE =

3.01 cm, 0.20 m^3 and 0.61 m; 16.79% , 47.39% and 15.83% of the mean variables, respectively). The model used to estimate the total volume per plot also performed reasonably well ($R^2 = 0.41$, $GR^2 = 0.31$, and $RMSE = 9.00$ m^3 ; 36.02% of the mean total volume per plot). The model used to estimate the number of trees per plot yielded the poorest fit ($R^2 = 0.22$, $GR^2 = 0.04$, with

$RMSE = 51.01$; 64.14% of the mean number of trees per plot) and the model used to estimate the mean height also provided a poor fit to the training data set ($R^2 = 0.25$, $GR^2 = 0.21$ and $RMSE = 2.73$ m; 24.96% of the mean height). Thus a large difference between GR^2 and R^2 indicates a severely over-parametrized model, i.e., a model that may show a good fit to the training dataset, but would not be generally applicable.

The predictor variables (spectral bands and indices and topographic variables) used to construct the basic functions can be evaluated after fitting a MARS model to the different dasometric variables. The parameter estimates of the MARS models fitted to the complete database are shown in Tab. 5.

The number of the predictor variables and the percentage contribution to the predictive power of the models varied widely in the models fitted (Fig. 2). For instance, the following variables were the most important (100% of GCV) in some models: slope (SLOPE), for estimating number of trees per plot; soil adjusted vegetation index (SAVI), for estimating mean diameter at base per plot, mean total tree height, mean crown diameter and mean volume; and the green band (B3) and short wave infrared 2 band (B7) for estimating the mean diameter at breast height and the total volume per plot. Likewise, the following variables contributed least to the predictive ability of the models (<25% of GCV): fraction of photosynthetically active radiation (FPAR), for estimating number of trees per plot; albedo (ALB) for estimating the mean diameter at base and the total volume per plot; the blue band (B2), for estimating the mean diameter at breast height, and DEM, wetness index (WETIND) and the red band (B4), for estimating the mean volume.

Fig. 3 shows the variation in k -fold cross-validation ($k = 2$ to 10), which is used to visualize problems such as overfitting. And finally, the spatial distribution of the daso-

Tab. 5 - Models obtained for the total database by MARS technique. (N): number of trees per plot; (DB mean): mean diameter at base (0.30 m, in cm); (DBH mean): mean diameter at breast height (1.30 m, in cm); (H mean): mean height (m); (CD mean): mean of the crown diameter (m); (V mean): mean volume (m^3); (TV): total volume (m^3) per plot.

Variable	MARS models
N	$+101.0029 - 0.2084 \cdot \max(0, 1.479 - B5) + 0.19820 \cdot \max(0, 0.695 - FPAR) - 2.5521 \cdot \max(0, ALB - 73) + 11.676 \cdot \max(0, 0.5495 - SLOPE)$
BD mean	$+23.7517 + 0.0926 \cdot \max(0, 0.310 - B3) - 0.0418 \cdot \max(0, B3 - 310) + 0.1064 \cdot \max(0, 0.283 - SAVI) + 0.1814 \cdot \max(0, SAVI - 250) - 0.3544 \cdot \max(0, 0.97 - ALB)$
DBH mean	$+28.5076 - 0.2359 \cdot \max(0, 0.105 - B2) + 0.1826 \cdot \max(0, B3 - 310) + 0.1356 \cdot \max(0, B4 - 354) + 0.9940 \cdot \max(0, SAVI - 208) - 0.2823 \cdot \max(0, SAVI - 214) + 1.9544 \cdot \max(0, SAVI - 218) - 2.9222 \cdot \max(0, ALB - 67) + 3.2205 \cdot \max(0, ALB - 70) - 0.2655 \cdot \max(0, ALB - 97) - 0.0347 \cdot \max(0, 0.2646 - DEM) - 0.0319 \cdot \max(0, DEM - 2646) - 7.2031 \cdot \max(0, CURV + 0.463) - 29.5239 \cdot \max(0, CURV - 0) + 41.0845 \cdot \max(0, PLANCURV - 0.107) + 58.5195 \cdot \max(0, -0.092 - PROF CURV) - 1.2622 \cdot \max(0, WETIND - 8.222)$
H mean	$+10.0624 + 0.0950 \cdot \max(0, SAVI - 238)$
CD mean	$+3.4718 - 0.0220 \cdot \max(0, B3 - 348) + 0.0210 \cdot \max(0, B4 - 387) + 0.0083 \cdot \max(0, 0.1247 - B5) + 0.0122 \cdot \max(0, B7 - 730) + 0.0255 \cdot \max(0, SAVI - 238) - 0.1455 \cdot \max(0, ALB - 97) + 2.100 \cdot \max(0, -0.185 - PROF CURV) - 0.1352 \cdot \max(0, 0.5495 - SLOPE)$
V mean	$+0.1153 + 0.0061 \cdot \max(0, 0.280 - B3) - 0.0060 \cdot \max(0, B3 - 289) + 0.0045 \cdot \max(0, B4 - 236) + 0.0174 \cdot \max(0, SAVI - 255) - 0.0123 \cdot \max(0, DEM - 2796) - 0.4713 \cdot \max(0, CURV - 0) + 1.9043 \cdot \max(0, -0.1851 - PROF CURV) - 0.1027 \cdot \max(0, WETIND - 10.51)$
TV	$+29.4140 - 0.0513 \cdot \max(0, B7 - 474) + 0.4461 \cdot \max(0, SAVI - 266) + 0.4821 \cdot \max(0, ALB - 97)$

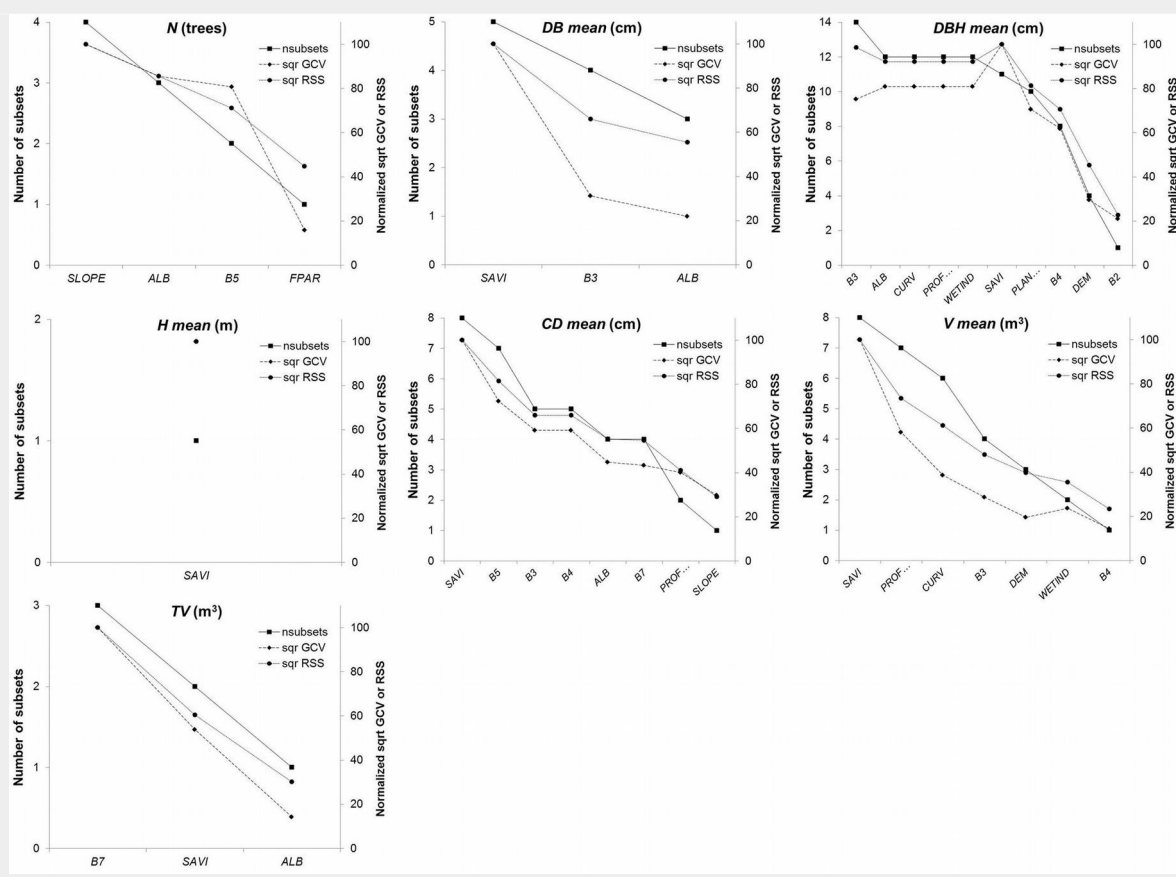


Fig. 2 - Importance and selection of predictor variables in the multivariate adaptive regression splines (MARS) models for each radiometric correction algorithm considered.

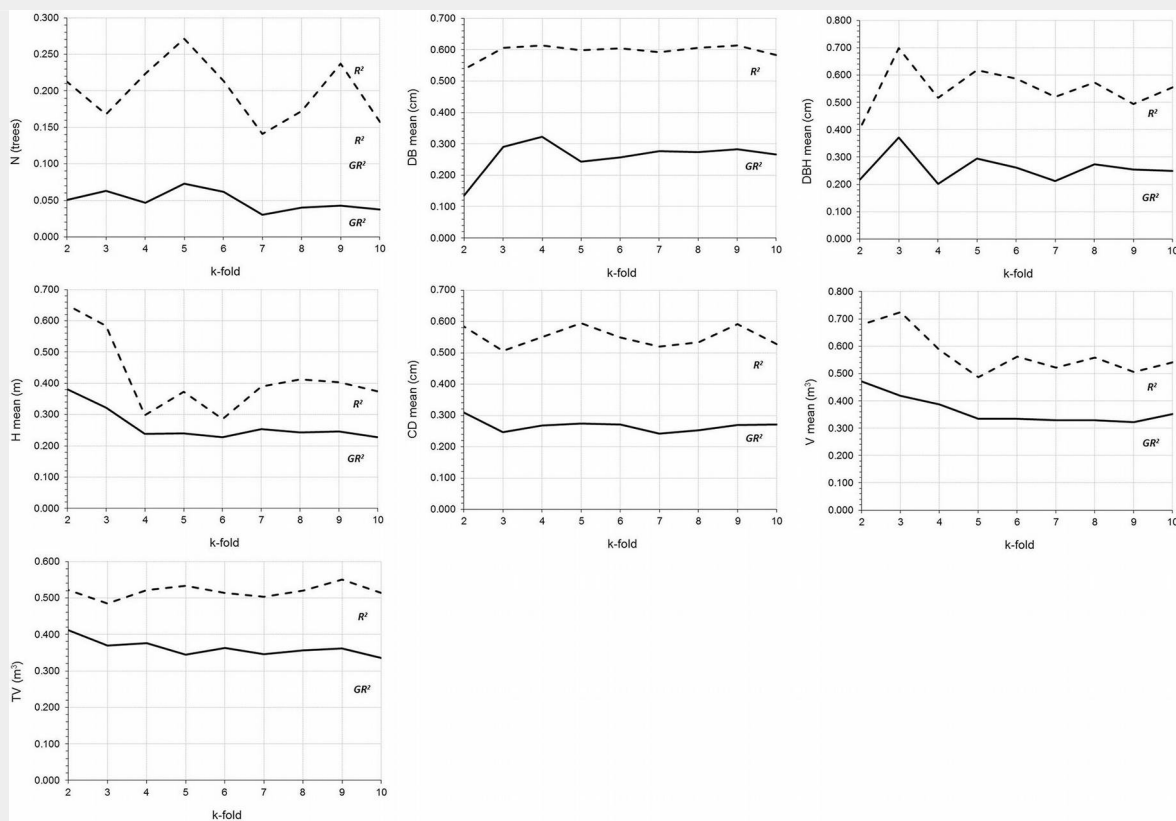


Fig. 3 - Variations in the goodness of fit statistics obtained by applying the cross-validation technique (nfolds = 2-10) to the MARS models.

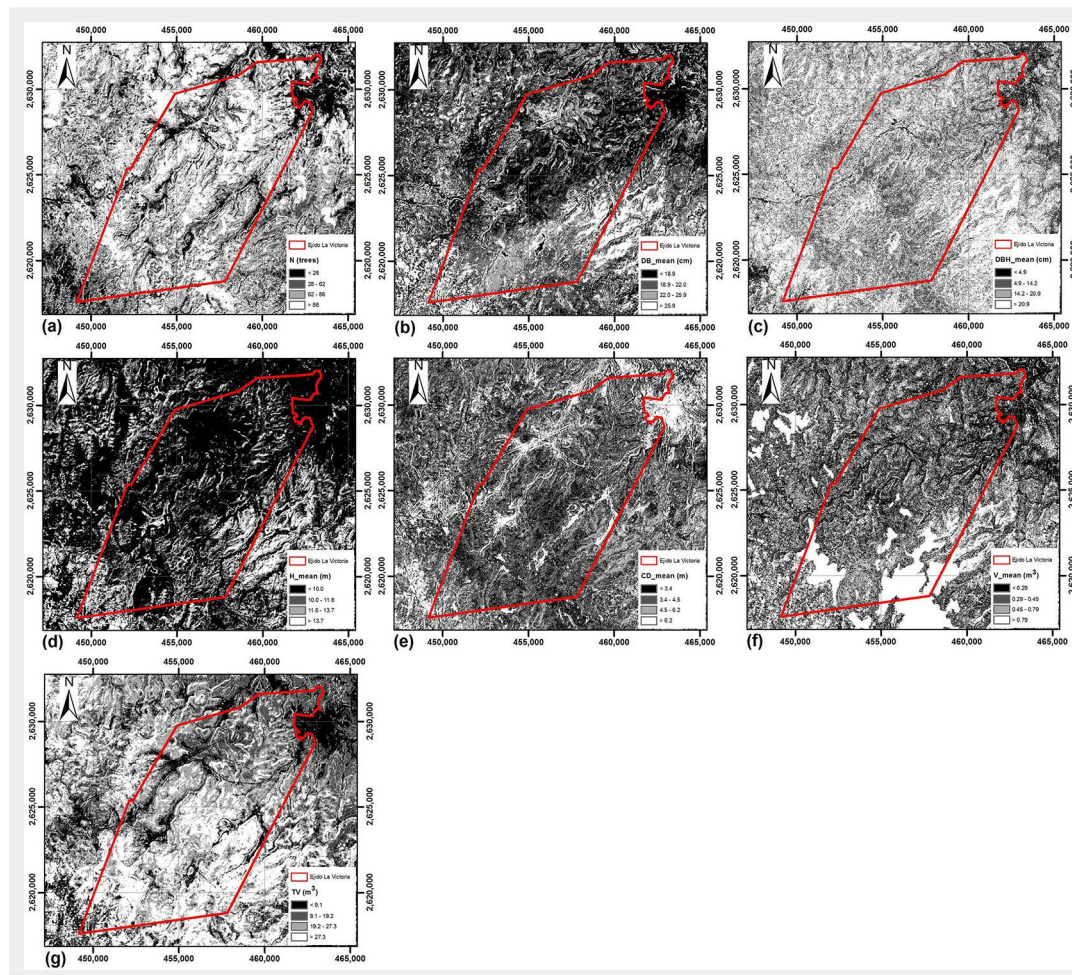


Fig. 4 - Spatial distribution of the stand variables in the study area, obtained by application of the MARS technique.

metric variables in the study area obtained by application of the MARS models (Tab. 5) is shown in Fig. 4.

Discussion

Predictive modeling is not an exact science (Frescino & Gretchen 2002), and model performance can be influenced by the method of statistical analysis, topographical factors (slope, exposure and elevation) and forest structure (proportion of young and mature trees, species and type of silvicultural management). The results obtained in this and other studies (Moisen & Frescino 2002, Zheng et al. 2009, Aertsen et al. 2010) were based on nonparametric techniques. In this case, the MARS technique yielded significantly more accurate predictions than the MLR excepting the variable *H mean* as a consequence of the smaller number of predictors (1 of 21) in the MARS model.

However, the use of very high resolution sensors can improve prediction of forest structural parameters by linear regression analysis, as reported by Ozdemir & Karnieli (2011). These authors used multispectral data (WorldView-2) and a stepwise regression technique to predict the number of trees per hectare ($R^2 = 0.38$, $RMSE = 109.56$ tree ha^{-1} ; 28.45%) and total volume per plot ($R^2 = 0.42$, $RMSE = 27.18$ m^3 ha^{-1} ; 44.20%) in a plantation forest in Israel.

Trotter et al. (1997) obtained high good-

ness-of-fit values ($R^2 \geq 0.85$ and $RMSE = 39.00$ m^3 ha^{-1}) for band 4, and for bands 3 and 4 of Landsat TM satellite images for estimating timber volume by simple and multiple linear regression methods, respectively. In a study monitoring growth of a young forest plantation in UK, they used generalized linear models (GLM) with logarithmic links and normal errors, and obtained higher values ($R^2 = 0.85$ and $RMSE = 0.88$ m; 28.38%) for the mean height, and lower values for the diameter at breast height ($R^2 = 0.36$ and $RMSE = 2.15$ cm; 27.24%) and for density ($R^2 = 0.03$ and $RMSE = 3.260.59$ tree ha^{-1} ; 98.62%) than those obtained in the present study. The differences in the results can be explained by the more heterogeneous conditions of the native vegetation in the present study area and, in the case of height, the higher R^2 values obtained in UK may also be related to the larger latitude in this country, which implies longer shadows at the time of the Landsat overpass, making it easier to estimate tree heights (Donoghue et al. 2004). In a study modeling forest stand structure attributes by multivariate regression and discriminant function analysis, Hall et al. (2006) obtained higher error values for height and crown cover ($R^2 = 0.57$ and 0.65 , and $RMSE = 12.0$ and 2.8 m, respectively). Coulston et al. (2012) estimated crown cover in forests in Utah and Michigan (USA) by using a beta regression model

and the random forest algorithm, and obtained also high error values ($R^2 = 0.65$ and 0.69 and $RMSE = 0.87$ and 0.89 m, respectively).

Cruz-Leyva et al. (2010) modeled even-aged *Pinus patula* and *P. teocote* forest stand variables in Hidalgo, Mexico by GIS variables and reflectance data derived from a multispectral SPOT 5 image and multiple linear regression. The volume was significantly related to tree crown cover, leaf area index (LAI), slope, elevation, mean annual temperature, maximum temperature of warmest month, mean annual precipitation, and precipitation of the wettest quarter ($R^2 = 0.72$ and 0.88 , $p < 0.001$).

Li et al. (2011) used stepwise linear regression and regression tree analysis, together with cross validation methods, to estimate height in young forests and obtained the following values for the first and second modeling techniques, respectively: $R^2 = 0.15$ and 0.19 and $RMSD$ (root mean square of the difference) = 10.32 and 6.07 (m) for the first group of variables by combining bands 1-5 and 7 to obtain the integrated forest index (IFI), a measure of the probability that a pixel corresponds to forest, the normalized difference vegetation index (NDVI) and the normalized burn ratio index (NBRI). The same authors obtained $R^2 = 0.88$ and 0.89 and $RMSD = 2.42$ and 2.13 (m) for the second group of variables (age since distur-

bance, derived from Vegetation Change Tracker [VCT] and Landsat time series stacks [LTSS] disturbance year map), and $R^2 = 0.69$ and 0.68 with $RMSD = 3.85$ and 3.73 (m) successively for the third group of variables (combining accumulated indexes: *FI*, *NDVI* and *NBRI*).

The null contribution of the *NDVI* (Fig. 2) can be attributed to the significant association between this index and both the leaf area index and the fraction of photosynthetically active radiation intercepted (Law & Waring 1994). The values of these variables are low in winter, which is when the satellite image was captured and the field sampling was carried out for the present study.

Some estimates are difficult to compare, because of differences in the sources of data, estimation methods used, period of data compilation and large differences between the study areas (Zhang et al. 2014).

The top performing model (R^2) was that used to estimate the mean diameter at breast height, probably because this model included the largest number of independent variables (Tab. 4), indicating problems of overfitting (GR^2) in the training data set (Fig. 3). The remaining unexplained variance in the spatial patterns may be related to the high heterogeneity of the dependent variables (García-Martín et al. 2006), due to the uneven-aged and tree species-rich structure of the Pueblo Nuevo forests in Durango. We therefore recommend increasing the number of observations in order to decrease the possibility of occurrence of this effect.

Considering the advantages and limitations of different remote sensing images, the medium-resolution (pixel size, 30 m) Landsat series is one of the most widely used for estimating dasometric variables (Agarwal et al. 2014, Pflugmacher et al. 2014, Dube & Mutanga 2015, Zhu & Liu 2015). The advantages of using the Landsat series are that numerous historical spatiotemporal archives are available and that the sensor is cheaper and more accessible than high resolution sensors, particularly for analysis of large areas (Wu et al. 2016).

Conclusions

We conclude that the spectral bands of the Landsat-8 OLI sensor can be used to estimate four of the seven stand variables (mean diameter at breast height, mean crown diameter, mean volume and total volume per plot) in the study region. Furthermore, the inclusion of aggregated values of vegetation indexes and terrain variables derived from the DEM contributed significantly to increasing the goodness-of-fit of the models.

Multivariate adaptive regression splines (MARS) models are more flexible than linear regression models. Application of the MARS technique facilitated handling and interpretation of large amounts of data representing complex structures. The predictors thus obtained for the study area

were accurate and the method proved suitable for modeling a large number of predictive variables.

References

- Aertsen W, Kint V, Orshoven JV, Ozkan K, Muys B (2010). Comparison and ranking of different modeling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modeling* 221: 1119-1130. - doi: [10.1016/j.ecolmodel.2010.01.007](https://doi.org/10.1016/j.ecolmodel.2010.01.007)
- Agarwal R, Ranjan P, Chipman H (2014). A new Bayesian ensemble of trees approach for land cover classification of satellite imagery. *Canadian Journal of Remote Sensing* 39 (6): 507-520. - doi: [10.5589/m14-003](https://doi.org/10.5589/m14-003)
- Alonso-Fernández JR, García-Nieto PJ, Díaz Muñiza C, Antón JC (2014). Modelling eutrophication and risk prevention in a reservoir in the Northwest of Spain by using multivariate adaptive regression splines analysis. *Ecological Engineering* 68: 80-89. - doi: [10.1016/j.ecoleng.2014.03.094](https://doi.org/10.1016/j.ecoleng.2014.03.094)
- 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)
- Baret F, Guyot G (1991). Potentials and limits of vegetation indices for LAI and PAR assessment. *Remote Sensing of Environment* 35 (2-3): 161-173. - doi: [10.1016/0034-4257\(91\)90009-U](https://doi.org/10.1016/0034-4257(91)90009-U)
- Brutsaert W (1975). On a derivable formula for long-wave radiation from clear skies. *Water Resources Research* 11 (5): 742-744. - doi: [10.1029/WR011i005p00742](https://doi.org/10.1029/WR011i005p00742)
- Corral-Rivas JJ, Diéguez-Aranda U, Corral-Rivas S, Castedo-Dorado F (2007). A merchantable volume system for major pine species in El Salto, Durango (Mexico). *Forest Ecology and Management* 238 (2007) 118-129. - doi: [10.1016/j.foreco.2006.09.074](https://doi.org/10.1016/j.foreco.2006.09.074)
- Coulston JW, Moisen GG, Wilson BT, Finco MV, Cohen WB, Brewer CK (2012). Modeling percent tree canopy cover: a pilot study. *Photogrammetric Engineering and Remote Sensing* 78 (7): 715-727. - doi: [10.14358/PERS.78.7.715](https://doi.org/10.14358/PERS.78.7.715)
- Cruz-Leyva IA, Valdez-Lazalde JR, Pérez G, Santos-Posadas HM (2010). Modelación espacial de área basal y volumen de madera en bosques manejados de *Pinus patula* y *P. teocote* en el ejido Atopixco, Hidalgo [Spatial modeling of basal area and tree volume in managed *Pinus patula* and *P. teocote* forests in the ejido Atopixco, Hidalgo]. *Madera y Bosques* 16 (3): 75-97. [in Spanish] [online] URL: http://www.scielo.org.mx/scielo.php?pid=S1405-04712010000300006&script=sci_arttext
- Donoghue DNM, Watt PJ, Dunford RW, Wilson J, Staples S, Smith S, Batts A, Wooding MJ (2004). An evaluation of the use of satellite data for monitoring *Picea sitchensis* plantation forest establishment and growth. *Forestry* 7 (5): 384-396. - doi: [10.1093/forestry/77.5.383](https://doi.org/10.1093/forestry/77.5.383)
- Dube T, Mutanga O (2015). Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying above-ground biomass in Mgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing* 101: 36-46. - doi: [10.1016/j.isprsjprs.2014.11.001](https://doi.org/10.1016/j.isprsjprs.2014.11.001)
- 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>
- Fassnacht FE, Hartig F, Latifi H, Berger C, Hernández J, Corvalán V, Koch B (2014). Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sensing of Environment* 154: 102-114. - doi: [10.1016/j.rse.2014.07.028](https://doi.org/10.1016/j.rse.2014.07.028)
- Franklin S, Hall R, Smith L, Gerylo G (2003). Discrimination of conifer height, age and crown closure classes using Landsat-5 TM imagery in the Canadian Northwest Territories. *International Journal of Remote Sensing* 24 (9): 1823-1834. - doi: [10.1080/01431160210144589](https://doi.org/10.1080/01431160210144589)
- Frescino TS, Gretchen GM (2002). Predictive mapping of forest attributes on the Fishlake National Forest. In: *Proceedings of the "4th Annual Forest Inventory and Analysis Symposium"*. New Orleans (LA, USA) 19-21 Nov 2002. Gen. Tech. Rep. NC-252, North Central Research Station, USDA Forest Service, St. Paul, MN, USA, pp. 257. [online] URL: http://www.fs.fed.us/rm/ogden/research/publications/downloads/proceedings2005_fiasym_frescino.pdf
- Friedman JH (1991). Multivariate adaptive regression splines. *The Annals of Statistics* 19 (1): 1-141. - doi: [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963)
- Gadow K, Sanchez S, Aguirre OA (2004). Manejo forestal con base científica [Science-based forest management]. *Madera y Bosques* 10 (2): 3-16. [in Spanish]
- García-Martín A, Pérez-Cabello F, Riva-Fernández J (2006). Evaluación de los recursos de biomasa residual forestal mediante imágenes del satélite Landsat y SIG [Resource assessment of residual forest biomass by Landsat satellite images and GIS]. *GeoFocus* 6: 205-230. [in Spanish] [online] URL: http://geofocus.rediris.es/2006/Articulo9_2006.pdf
- Gleason CJ, Im J (2011). A review of remote sensing of forest biomass and biofuel: options for small-area applications. *Giscience and Remote Sensing* 48 (2): 141-170. - doi: [10.2747/1548-1603.48.2.141](https://doi.org/10.2747/1548-1603.48.2.141)
- González-Elizondo MS, González-Elizondo M, Tena-Flores JA, Ruacho-González L, López-Enríquez L (2012). Vegetación de la Sierra Madre Occidental, México: una síntesis [Vegetation of the Sierra Madre Occidental, Mexico: a synthesis]. *Acta Botánica Mexicana* 100: 351-403. [in Spanish] [online] URL: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=So187-71512012000300012
- Hall RJ, Skakun RS, Arsenault EJ, Case BS (2006). Modeling forest stand structure attributes using Landsat ETM+ data: application to mapping of aboveground biomass and stand volume. *Forest Ecology and Management* 225 (1): 378-390. - doi: [10.1016/j.foreco.2006.01.014](https://doi.org/10.1016/j.foreco.2006.01.014)
- Hastie T, Tibshirani R, Friedman J (2009). *The*

- elements of statistical learning: data mining, inference, and prediction (2nd edn). Springer-Verlag, New York, USA, pp. 745.
- Huete AR (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25 (3): 295-309. - doi: [10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- INEGI (2014). Continuo de elevaciones mexicano 3.0 (CEM 3.0) [Continuous elevation 3.0 (CEM 3.0)]. Instituto Nacional de Estadística Geográfica e Informática, DF, México. [online] URL: <http://www.inegi.org.mx/geo/contenidos/datos/relieve/continental/descarga.aspx>
- Ji L, Wylie BK, Noss DR, Peterson B, Waldrop MP, McFarland JW, Rover J, Hollingsworth TN (2012). Estimating aboveground biomass in interior Alaska with Landsat data and field measurements. *International Journal of Applied Earth Observation and Geoinformation* 18: 451-461. - doi: [10.1016/j.jag.2012.03.019](https://doi.org/10.1016/j.jag.2012.03.019)
- Law B, Waring R (1994). Remote sensing of leaf area index and radiation intercepted by understory vegetation. *Ecological Applications* 4 (2): 272-279. - doi: [10.2307/1941933](https://doi.org/10.2307/1941933)
- Li A, Huang C, Sun G, Shi H, Toney C, Zhu Z, Rollins MG, Goward SN, Masek JG (2011). Modeling the height of young forests regenerating from recent disturbances in Mississippi using Landsat and ICESat data. *Remote Sensing of Environment* 115 (8): 1837-1849. - doi: [10.1016/j.rse.2011.03.001](https://doi.org/10.1016/j.rse.2011.03.001)
- López-Serrano PM, López Sánchez CA, Solís-Moreno R, Corral-Rivas JJ (2016). Geospatial Estimation of above Ground Forest Biomass in the Sierra Madre Occidental in the State of Durango, Mexico. *Forests* 7: 70. - doi: [10.3390/f7030070](https://doi.org/10.3390/f7030070)
- Luther JE, Fournier RA, Piercey DE, Guindon L, Hall RJ (2006). Biomass mapping using forest type and structure derived from Landsat TM imagery. *International Journal of Applied Earth Observation and Geoinformation* 8 (3): 173-187. - doi: [10.1016/j.jag.2005.09.002](https://doi.org/10.1016/j.jag.2005.09.002)
- 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/1989/0000035/00000001/art00010>
- Martínez-Antúnez P, Wehenkel C, Hernández-Díaz JC, Corral-Rivas JJ (2015). Use of the Weibull function to model maximum probability of abundance of tree species in northwest Mexico. *Annals of Forest Science* 72 (2): 243-251. - doi: [10.1007/s13595-014-0420-2](https://doi.org/10.1007/s13595-014-0420-2)
- Moisen GG, Frescino TS (2002). Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling* 157: 209-225. - doi: [10.1016/S0304-3800\(02\)00197-7](https://doi.org/10.1016/S0304-3800(02)00197-7)
- 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.
- Ozdemir I, Karnieli A (2011). Predicting forest structural parameters using the image texture derived from WorldView-2 multispectral imagery in a dryland forest, Israel. *International Journal of Applied Earth Observation and Geoinformation* 13: 701-710. - doi: [10.1016/j.jag.2011.05.006](https://doi.org/10.1016/j.jag.2011.05.006)
- Poulain M, Peña M, Schmidt A, Schmidt H, Schulte A (2010). Relationships between forest variables and remote sensing data in a *Nothofagus pumilio* forest. *Geocarto International* 25 (1): 25-43. - doi: [10.1080/10106040902803558](https://doi.org/10.1080/10106040902803558)
- Pflugmacher D, Cohen WB, Kennedy RE, Yang Z (2014). Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics. *Remote Sensing of Environment* 151: 124-137. - doi: [10.1016/j.rse.2013.05.033](https://doi.org/10.1016/j.rse.2013.05.033)
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, pp. 3551. [online] URL: <http://www.R-project.org/>
- 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/1SRNyMA-CONAFOR> (2007). Plan estratégico forestal 2030 [Strategic forest plan 2030]. Gobierno del estado de Durango, Durango, México, pp. 242. [in Spanish]
- Trotter CM, Dymond JR, Goulding CJ (1997). Estimation of timber volume in a coniferous plantation forest using Landsat TM. *International Journal Remote Sensing* 18: 2209-2223. - doi: [10.1080/014311697217846](https://doi.org/10.1080/014311697217846)
- Venables WN, Ripley BD (2002). Modern applied statistics with S (4th edn). Springer, New York, USA, pp. 498.
- Vidoli F (2011). Evaluating the water sector in Italy through a two stage method using the conditional robust nonparametric frontier and multivariate adaptive regression splines. *European Journal of Operational Research* 212: 583-595. - doi: [10.1016/j.ejor.2011.02.003](https://doi.org/10.1016/j.ejor.2011.02.003)
- Wehenkel C, Corral-Rivas JJ, Hernández-Díaz JC, Gadov KV (2011). Estimating balanced structure areas in multi-species forests on the Sierra Madre Occidental, Mexico. *Annals of Forest Science* 68: 385-394. - doi: [10.1007/s13595-011-0027-9](https://doi.org/10.1007/s13595-011-0027-9)
- 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/>
- Wu C, Shen H, Wang K, Shen A, Deng J, Gan M (2016). Landsat imagery-based above ground biomass estimation and change investigation related to human activities. *Sustainability* 8 (2): 159. - doi: [10.3390/su8020159](https://doi.org/10.3390/su8020159)
- 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)
- Zhang J, Huang S, Hogg EH, Liefers V, Qin Y, He F (2014). Estimating spatial variation in Alberta forest biomass from a combination of forest inventory and remote sensing data. *Biogeosciences* 11: 2793-2808. - doi: [10.5194/bg-11-2793-2014](https://doi.org/10.5194/bg-11-2793-2014)
- Zheng HL, Chen X, Han X, Zhao X, Ma Y (2009). Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: the importance of phosphorus application rates under drought conditions. *Agriculture, Ecosystems and Environment* 132: 98-105. - doi: [10.1016/j.agee.2009.03.004](https://doi.org/10.1016/j.agee.2009.03.004)
- Zhu X, Liu D (2015). Improving forest above-ground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS Journal of Photogrammetry and Remote Sensing* 102: 222-231. - doi: [10.1016/j.isprsjprs.2014.08.014](https://doi.org/10.1016/j.isprsjprs.2014.08.014)

Supplementary Material

Tab. S1 - Vegetation indexes and other derived parameters generated from the spectral bands of the sensor.

Tab. S2 - Terrain attributes derived from the digital elevation model (DEM).

Link: Lopez-Sanchez_1891@suppl001.pdf