UNIVERSIDAD JUÁREZ DEL ESTADO DE DURANGO

DOCTORADO INSTITUCIONAL EN CIENCIAS AGROPECUARIAS Y FORESTALES

Estimación de la biomasa aérea forestal a nivel rodal mediante sensores remotos pasivos en el estado de Durango

Por:
Pablito Marcelo López Serrano

Como requisito parcial para obtener el grado de

DOCTOR EN CIENCIAS AGROPECUARIAS Y FORESTALES

Opción Terminal: Manejo, Aprovechamiento y Conservación de Recursos Naturales

Director de Tesis: Dr. Carlos Antonio López Sánchez

Durango, Dgo., México

2016

Universidad Juárez del Estado de Durango

Doctorado Institucional en Ciencias Agropecuarias y Forestales

Los abajo firmantes, certifican que la tesis de doctorado que se presenta como requisito parcial para la obtención del Grado de Doctor en Ciencias Agropecuarias y Forestales por parte de

Pablito Marcelo López Serrano

ha cumplido con los requisitos estipulados por la UJED

_____
Dr. Carlos Antonio López Sánchez
Director de Tesis

_____
Dr. Juan Gabriel Álvarez González
Co-Director de Tesis

_____
Dr. José Javier Corral Rivas
Asesor

_____
Dr. Raúl Solís Moreno
Asesor

_____
Dr. Ramón Díaz Varela
Asesor Externo

DEDICATORIA

A mis padres:

Jacinto López Ramírez (✝)

Paula Serrano Pereda (✝)

"Quienes me mostraron que la consumación plena de mi existencia se logra con la fe puesta en lo que soy, puedo y hago".

A mi familia:

"Como testimonio de mi gratitud ilimitada;

A mi hija Emily Fernanda López Saucedo, por que su presencia ha sido y sera siempre el motivo mas grande que me han impulsado para lograr mis metas;

A mi esposa Norma A. Saucedo Ibarra por su gran comprensión, amor y tolerancia a lo largo de este camino que se llama vida".

A la familia Saucedo Ibarra:

"En especial a la Sra M. de Lourdes, por su apoyo incondicional, en los ultimos 15 años".

AGRADECIMIENTOS

CONTENIDO

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

RESUMEN

La estimación de la biomasa forestal aérea ("Above Ground Biomass" AGB) generalmente suele realizarse mediante un método directo (destructivo), que resulta ser preciso pero conlleva altos costos económicos y en tiempo y debe ser realizado en áreas pequeñas. Las imágenes de satélite permiten registrar información de grandes áreas de terreno como la Sierra Madre Occidental en el Estado de Durango, la aplicación a esta información de los nuevos métodos de estadística paramétrica y no paramétrica, permiten la cuantificación de un gran número de variables de interés forestal con suficiente precisión, y entre ellas se encuentra la biomasa forestal aérea. El estudio se desarrolló en la Sierra Madre Occidental, en el Estado de Durango, abarcando una superficie de 6.33 millones de hectáreas aproximadamente. Se establecieron Sitios Permanentes de Investigación Forestal y Suelos (SPIFyS) de 2500 m^2 de superficie. Se utilizaron imágenes del sensor Landsat-5 TM tratadas con diferentes algoritmos de corrección radiométrica, variables de textura derivadas del NDVI, asi como variables topográficas obtenidas del modelo digital de elevación. El análisis estadístico fue realizado mediante técnicas paramétricas y no paramétricas. Las variables espectrales de mayor importancia para cuantificar la AGB fueron las del espectro infrarrojo (cercano y medio). Por otra parte, el índice de humedad, como variable topográfica, presentó una fuerte asociación con la presencia de biomasa. Las técnicas estadísticas no paramétricas fueron las que mejor predijeron la AGB. Estos resultados demuestran que el uso de sensores remotos representan una alternativa importante con bajos costos (económico y de tiempo), en la cuantificación y monitoreo de AGB en la Sierra Madre Occidental de Durango, lo que favorecerá la toma de decisiones para el manejo de los recursos forestales.

Palabras clave: AGB, Landsat-5 TM, SPIFyS, no paramétrica

SUMMARY

The estimation of aerial forest biomass ("Above Ground Biomass" AGB) usually is usually performed by a direct method (destructive), which turns out to be accurate but carries high economic costs and time and should be done in small areas. Satellite images allow information recording large areas of land as the Sierra Madre Occidental in the State of Durango, applying this information to new methods of parametric and non-parametric statistics, allow quantification of a large number of variables of interest with sufficient precision forestry, and including aerial forest biomass is located. The study was conducted in the Sierra Madre Occidental, in the State of Durango, covering an area of approximately 6.33 million hectares. Permanent sites Forest and Soil Research (SPIFyS) of 2500 m^2 were established. Topographic variables derived from digital elevation model Landsat-5 TM sensor treated with different radiometric correction algorithms, texture variables derived from NDVI were used, as well. Statistical analysis was performed using parametric and nonparametric techniques. The spectral variables most important to quantify the AGB were the infrared spectrum (near and medium). Moreover, the moisture content, such as topographical variable, showed a strong association with the presence of biomass. Nonparametric statistical techniques were those that best predicted the AGB. These results demonstrate that the use of remote sensors represent an important Altena low (financial and time) costs, quantification and monitoring of AGB in the Sierra Madre Occidental in Durango, which will facilitate decision making for the management of forest resources.

Keys words: AGB, Landsat-5 TM, SPIFyS, no paramétric

CAPÍTULO 1.

ORGANIZACIÓN DE LA TESIS

El presente documento está dividido en seis capítulos. El primero se corresponde con la introducción general. El segundo, tercero, cuarto y quinto están conformados por los artículos científicos publicados en revistas indexadas y derivados del trabajo de tesis doctoral. En el capítulo sexto se ubican las conclusiones generales.

Los artículos que conforman los capítulos del segundo al quinto, a la fecha de redacción de este documento, han sido aceptados y publicados en las revistas: *iForest. vol. 9, pp. 226-234. DOI:10.3832/ifor1504-008, Forests 2016, 7(3), 70, DOI:10.3390/f7030070, Canadian Journal of Remote Sensing,* (en prensa) *y Remote sensing, 8(5), 369, DOI:10.3390/rs8050369,* respectivamente. A continuación se hace una breve descripción de cada capitulo:

Capítulo 1: Introducción general. Aporta un panorama general del documento de tesis.

Capítulo 2: "Estimación de la biomasa de bosques mixtos y de diferentes edades usando datos espectrales y un modelo híbrido que combina árboles de regresión y modelos lineales", cuyo título publicado en inglés es *"Estimating biomass of mixed and uneven-aged forests using spectral data and a hybrid model combining regression trees and linear models".* El objetivo del presente estudio fue modelar la biomasa forestal en bosques mixtos de edad diferente en la Sierra Madre Occidental (SMO), mediante el uso de imágenes Landsat-5 TM, variables del terreno y datos de los inventarios forestales tradicionales (terrestres) obtenidos a partir de una red de parcelas permanentes de muestreo (SPIFyS). Se compararon dos enfoques distintos: el método de modelado habitual de ajuste de una relación lineal entre la biomasa obtenida en sitios de campo e imágenes de satelite, y un nuevo enfoque que consiste en un modelo híbrido que combina árboles de regresión y modelos lineales en los nodos finales del árbol. Por lo que sabemos, esta es la primera vez que este enfoque híbrido se ha utilizado para modelar la biomasa con datos de teledetección.

Capítulo 3: "Estimacion geospacial de la biomasa forestal aérea en la Sierra Madre Occidental", cuyo título publicado en inglés es "*Geospatial estimation of above ground forest biomass in the Sierra Madre Occidental in the state of Durango, México*". El principal objetivo de este estudio fue estimar la AGB en la SMO mediante el análisis de datos

espectrales de media resolución bajo tratamiento de corrección de reflectancia aparente en el techo de la atmósfera (ToA's) y la reflectancia de la superficie (SR) usando datos de campo de una red de 201 sitios permanentes (SPIFyS), instalados por muestreo sistemático en la zona de estudio en el año 2011. Se utilizó el algoritmo M5P que construye modelos lineales basados en árboles de regresión y clasificación.

Capítulo 4: "Una comparación de técnicas de machine learning aplicando datos espectrales de Landsat-5 TM para la estimación de la biomasa", cuyo título en inglés es "*A comparison of machine learning techniques applied toLandsat-5 TM spectral data for biomass estimation*". El objetivo de este estudio fue comparar el desempeño de tres de las técnicas no paramétricas más comunes reportadas en la literatura (Support Vector Machine [SVM], k-Nearest Neighbours [kNN] y Random Forest [RF]) y la técnica paramétrica de regresión lineal multiple (MLR) para estimar AGB a partir de datos espectrales del sensor Landsat-5 TM, de parámetros de textura derivadas del índice de vegeacion de diferencia normalizada (NDVI) y de variables topográficas derivadas del modelo digital de elevación.

Capítulo 5: "Evaluación de algoritmos de corrección radiométrica en la estimación de biomasa aérea forestal mediante datos del sensor Landsat-5 TM", cuyo título en inglés es "*Evaluation of Radiometric and Atmospheric Correction Algorithms for Aboveground Forest Biomass Estimation Using Landsat 5 TM Data*". El objetivo del estudio fue evaluar cuatro algoritmos de corrección atmosférica (reflectancias en superficie) ATCOR, FLAASH, COST y 6S, junto con el algoritmo de corrección radiométrica ToA (reflectancia en el sensor), para estimar la AGB en los bosques templados al noreste del estado de Durango, México. Como fuente de datos se utilizó una imagen del sensor Landsat-5 TM y parámetros derivados del modelo digital de elevaciones (MDE) a partir de datos de campo procedentes de 99 sitios permanentes (SPIFyS) establecidos durante el invierno del año 2011 a través de un muestreo sistemático. La cuantificación de AGB se realizó mediante la técnica no paramétrica denominada "Multivariate Adaptative Regression Splines" (MARS).

Capítulo 6: Conclusiones generales. En este último capítulo se exponen los hallazgos más relevantes de los distintos resultados derivados de cada artículo en el trabajo de investigación doctoral y se hacen algunas recomendaciones para investigaciones futuras.

INTRODUCCION GENERAL

La Sierra Madre Occidental (SMO) es un área de gran interés ecológico por su alta heterogeneidad ambiental. El estudio de la biomasa forestal aérea (AGB) para este tipo de ecosistemas generalmente suele realizarse mediante un método directo (destructivo), que resulta ser preciso, pero conlleva altos costos, tiempo y son realizados en áreas pequeñas. En este sentido, la emergencia de la teledetección satelital cuantitativa con diferentes tipos de sensores (activos y pasivos), permite estimar parámetros dasométricos habitualmente medidos en inventarios forestales en menor costo, tiempo y abarcando grandes superficies, cuyos métodos son divididos en físicos y estadísticos. Asimismo el tratamiento de las imágenes de satélite es un importante paso para mejorar la calidad de los datos e interpretación, sobre todo para ecosistemas forestales, que generalmente se encuentran en superficies accidentadas. Dicho tratamiento consiste en corregir algunos efectos nebulosos ocacionados por la absorción o dispersión que provocan las partículas en suspensión (aerosoles) y otros elementos que integran la atmósfera. Asu vez, la generación de índices de vegetación y variables de textura derivadas de las imágenes satelitales se emplean como variables auxiliares en la estimación de la AGB.

Por otro lado, la utilización de métodos estadísticos para generar modelos potencialmente más complejos y precisos en la estimacion de la AGB, es uno de los temas mas importantes que se consideran al momento de trabajar con imágenes de satélite con respecto a los datos de campo. Es por ello, que la implementación de técnicas de machine learning, ofrece una gama de técnicas paramétricas y no paramétricas con mecanismos de validación cruzada para evaluar la calidad de una determinada técnica de ajuste y minimizar el riesgo de sobreajuste a los datos de entrenamiento.

En este sentido, los resultados de este documento de tesis permiten establecer diferentes líneas futuras para estimar la AGB mediante sensores remotos bajo diferentes técnicas estadísticas, asi como diferentes técnicas de correccion radiométrica usadas en sensores de media resolución, como Landsat 5 TM.

REVISION DE LITERATURA

Sensores remotos pasivos

La Sierra Madre Occidental (SMO) es una provincia de gran interés ecológico por su alta heterogeneidad ambiental, atribuido a la existencia de una gran diversidad fisiográfica y climática (González *et al.,* 2012). Aunado a ello, su importancia se incrementa al contar con especies de pinos y encinos, que son altamente representativos y de gran interés económico en los ecosistemas de México y del mundo (Sánchez *et al.*, 2003). Dentro de esta provincia ecológica se ubica el estado de Durango (71.5% de su superficie), que está considerado como la primer reserva nacional forestal, generando entre el 25 y 30% de la producción maderable nacional con un total de 5.5 millones de m^3 de madera en rollo anuales, además de ser una importante fuente de servicios ambientales (SEMARNAT, 2011). Uno de los principales recursos de estos ecosistemas es la biomasa forestal aérea ("Above Ground Biomass" AGB), que generalmente suele cuanatificarse mediante un método directo (destructivo), que resulta ser preciso pero conlleva altos costos económicos y en tiempo y que, por tanto, sólo son viables en áreas pequeñas (Ketterings *et al.,* 2001, Zianis and Mencuccini, 2004, Walker *et al.,* 2011). En este sentido, la emergencia de la teledetección satelital cuantitativa ha facilitado evitar estos inconvenientes, ya que permite estimar parámetros dasométricos habitualmente medidos en inventarios forestales en base a un menor costo, tiempo y abarcando grandes superficies (Liang, 2007). Un buen número de estudios han utilizado estas tecnologías emergentes (imágenes de satélite) para la estimación de la AGB en diferentes ecosistemas (por ejemplo, Muukkonen and Heiskanen, 2005, Fuchs, *et al.*, 2009 o Hernández-Stefanoni *et al.,* 2011).

Mineria de datos

Por otro lado, independientemente del tipo de sensor remoto utilizado para dichas estimaciones, la precisión y la estimación del error de los modelos obtenidos varían en función de una serie de factores, tales como la estructura de los datos de campo o la técnica estadística empleada (Ghosh *et al.,* 2014). Actualmente, las estimaciones espaciales de AGB se han centrado en el uso de algoritmos informáticos (machine learning) y técnicas no paramétricas

con el fin de construir modelos capaces de hacer predicciones a partir de una estructura compleja de datos (Lima *et al.,* 2013, Fassnacht *et al.,* 2014).

Las metodologías incluidas dentro de lo que se denomina "machine learning" comprenden un conjunto de técnicas estadísticas que utilizan un enfoque inductivo automático para reconocer patrones en unos datos de entrenamiento, una vez aprendido dicho patrón, se aplica al resto de los datos de validación para proporcionar una predicción, cuando la variable de interés es cuantitativa, o una clasificación cuando la variable de interés es cualitativa (Cracknell and Reading, 2014). En la última década, las técnicas no paramétricas más utilizadas en machine learning con el fin de desarrollar modelos predictivos de AGB en grandes áreas han sido: Support Vector Machines (SVM), k-Nearest Neighbour (k-NN) y Random Forest (RF) (Fassnacht *et al.*, 2014).

La aplicación de estas técnicas permite generar modelos potencialmente más complejos y precisos en comparación con la regresión lineal múltiple (MLR), que es la técnica paramétrica que tradicionalmente se ha venido aplicando en estos estudios (Morel *et al.,* 2012, Næsset *et al.,* 2013). Aunque el empleo de metodologías de machine learning puede mejorar los resultados de MLR, tienen el inconveniente de poder provocar un sobreajuste del modelo a los datos de entrenamiento (Hawkins, 2004). Este inconveniente debe ser tenido muy en cuenta en la estadística espacial porque puede llevar a una estimación sesgada de los parámetros del modelo y a reducir la exactitud de las predicciones (Johnson and Omland, 2004). Generalmente, las medidas más comunes para evaluar la bondad del ajuste de los modelos derivados de datos espectrales son el coeficiente de determinación (R^2) y la raíz del error cuadrático medio (*RMSE*) (Lumbres and Lee, 2014, Porter *et al.,* 2014). Cabe señalar que con estas medidas se reporta el desempeño del modelo predictivo de datos que se utilizaron para ajustar el modelo, sin embargo, al carecer de un análisis de validación independiente, el error en la aplicación práctica resulta ser más grande del que se reporta (Castillo-Santiago *et al.,* 2013, Fernández-Manso *et al.,* 2014). Es por ello la importancia de la aplicación de mecanismos como la validación cruzada ("cross-validation" CV), que permite evaluar la calidad de una determinada técnica de ajuste y minimizar el riesgo de sobreajuste a los datos de entrenamiento (Molinaro *et al.,* 2005). Pero aun así, cuando se realizan comparaciones entre los resultados de rendimiento de múltiples métodos a una misma base de datos, es

también necesario realizar un estudio estadístico que apoye y justifique las conclusiones alcanzadas (García *et al.,* 2010).

Algoritmos de correccion radiométrica

El tratamiento de las imágenes de satélite es un importante paso para mejorar la calidad de los datos y su interpretación, sobre todo para ecosistemas forestales, que generalmente se encuentran en superficies accidentadas (Richter, 2013) en donde el efecto de las condiciones topográficas causan variaciones en los valores de reflectancia en función de la geometría del terreno y del ángulo de elevación solar (Richter and Schläpfer, 2011). Además, la calidad de los productos finales obtenidos a partir de las imágenes depende, en gran medida, de una eficiente calibración de los parámetros del sensor y de la corrección de las potenciales alteraciones en la imagen debidas a efectos de distorsion radiométrica y de la atmósfera (Li *et al.,* 2010). Por consiguiente, la aplicación de una corrección topográfica y atmosférica se hace necesaria para contrarrestar dichos efectos, considerando atributos del modelo digital de elevación (MDE) como la pendiente, orientación, sombreado, "skyview" y altitud, junto con parámetros de calibración de los sensores y características de la atmosfera y radiación solar (Balthazar *et al.,* 2012, Richter, 2013). Estos parámetros primarios, al igual que los secundarios (índices topograficos), derivados estos últimos del MDE, están estrechamente relacionados con índices de diversidad de especies o estructura del bosque, habiéndose aplicado para describir procesos hidrológicos, geomorfológicos y ecológicos (Moore and Nieber ,1989, Wilson and Gallant, 2000).

Índices de Vegetación (IV)

El uso de índices de vegetación (IV) que resultan de las combinaciones entre bandas espectrales, como indicadores de parámetros biofísicos de la vegetación, tienen como objetivo resaltar la contribución de la vegetación activa en la respuesta espectral de una superficie, minimizando el efecto de otros factores, tales como suelo, restos de vegetación seca o atmósfera (Gilabert *et al.,* 1997). Estos IV están basados en el alto contraste existente entre diferentes bandas de las imágenes de satélite para diferentes parámetros de interés de la cobertura vegetal, como pueden ser la banda del rojo (R) y la del infrarrojo cercano (NIR), para discriminar la vegetación viva y verde (Tucker, 1979).

Variables de Textura de la imagen

Otro proceso que permite la obtención de variables auxiliares para la modelación con imágenes satélite es el análisis de textura, que es una cuantificación de la variación en el dominio espacial de valores en tonos de grises mediante una matriz de co-ocurrencia (GLCM, Grey Level), siendo muy útil a la hora de la identificación de áreas u objetos de interés dentro de la imagen (Haralick *et al*., 1973, Botero and Restrepo, 2010) y que han sido empleados en análisis clasificatorios y de modelización de características de la vegetación (Diaz-Varela *et al.,* 2011).

Variables de Terreno

Dado que la topografía constituye uno de los principales factores que regula la humedad del suelo, a partir de ella se han generado índices topográficos e hidrológicos para predecir la humedad del suelo en función del relieve (Beven and Kirkby, 1979, O´Loughlin, 1986, Barling *et al*. 1994). Además, los atributos del terreno muestran una relación directa con la presencia, desarrollo y evaluación de las especies forestales (McNab, 1989, Roberts and Cooper, 1989). Diversos autores coinciden en destacar la influencia de los distintos atributos topográficos en función del ambiente, principalmente, bajo estados secos o húmedos, reportando modelos de humedad que reflejan patrones en la vegetación (Schume *et al*., 2003, Hupet and Vanclooster, 2005), en la textura del suelo (Price and Bauer, 1984, Seyfried, 1998) y/o en características del uso de la tierra (Famiglietti *et al*., 1998, Qiu *et al*., 2001). Estas razones han hecho que los modelos digitales de elevación (MDE) constituyan una fuente de datos con alto potencial para caracterizar el relieve en forma cuantitativa, siendo un insumo importante para cuantificar la biomasa aérea forestal (Chen *et al*. 2004, Díaz-Varela *et al*. 2011).

Referencias

Balthazar, V., Veerle, V., Eric, F.L. 2012. Evaluation and parameterization of ATCOR3 topographic correction method for forest cover mapping in mountain areas. Int. J. Appl. Earth Obs. Geoinform.18, 436–450.

Barling, R., Moore, I., Grayson, R. 1994. A quasi-dynamic wetness index for characterizing the spatial distribution of zones of surface saturation and soil water content. Water Resources Research 30, 1029-1044.

Beven, K., Kirkby, M. 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Science Bulletin. 24, 43–69.

Botero V.J.S., Restrepo A. 2010. Análisis de textura en panes usando la matriz de coocurrencia. Revista Politécnica 6(10): 74-80.

Castillo-Santiago, M.A, Ghilardi, A., Oyama, K., Hermadez-Stefanoni, J.L., Torres, I., Flamenco-Sandoval, A., Fernández, A., Jean-Francois, M., 2013. Estimating the spatial distribution of Woody biomass suitable for charcoal making from remote sensing and geostatistics in central Mexico. Energy for Sustainable Development, Vol. 17, No. 2, pp. 177-188.

Chen, D., Stow. D.A., 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.

Cracknell, J.M., Reading, M.A. 2014. Reading. Geological mapping using remote sensing data: A comparison of ve machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Computers & Geosciences 63, 22-33.

Díaz-Varela, R.A., 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.

Famiglietti, J. S., Rudnicki, J.W., Rodell, M. 1998. Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas. Journal of Hydrology 210(1-4), 259-281.

Fassnacht, F.E., 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.

Fernandez-Manso, O., Fernandez-Manso, A., Quintano, C. 2014. Estimation of aboveground biomass in Mediterranean forests by statistical modelling of ASTER frection images. International Journal of Applied Earth Observation and Geoinformation. 31:45-56

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.

García, S., Fernández, A., Luengo, J., Herrera, F., 2010. Advanced non-parametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. Information Sciences, Vol. 180, No. 10, pp. 2044-2064.

Ghosh, A., Fassnacht, F.E., Joshi, P.K., Koch, B. 2014. A framework for map ping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. International Journal of Applied Earth Observation and Geoinformation 26, 49-63.

Gilabert, M.A., González-Piqueras, J. y García-Haro, J. 1997. Acerca de los índices de vegetación. Revista de Teledetección. 8. Pp. 35-45.

González Elizondo, M.S., González Elizondo, M., Tena Flores, J.A., Ruacho González, L., López-Enríquez, I.L. 2012. Vegetación de la Sierra Madre Occidental, México: una síntesis. Acta botánica Mexicana 100, 351-403.

Haralick RM, Shanmugam K, Dinstein I. 1973. Textural features for image classification. IEEE Transactions of Systems, Man, and Cybernetics 6: 610-621.

Hawkins, D. M. 2012. The Problem of Overfitting. J. Chem. Inf. Comput. Sci. 44, 1-12.

Hernández-Stefanoni, J.L., Gallardo-Cruz, J.A., Meave, J.A., Dupuy, J.M. 2011. Combining geostatistical models and remotely sensed data to improve tropical plant richness mapping. Ecological Indicators 11, 1046–1056.

Hupet, F., Vanclooster, M. 2005. Microvariability of hydrological processes at the maize row scale: implications for soil water content measurements and evaporations fluxes and evaporation estimates. Journal of Hydrology 303, 247-270.

Johnson, J.B., Omland, K.S. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution.19,101–108

Ketterings, Q.M., Coe, R., Van, M.N., Ambagau, Y., Palm, C.A. 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, 199–209.

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.

Liang, S. 2007. Recent developments in estimating land surface biogeophysical variables from optical remote sensing. Progress in Physical Geography 31, 501-516.

Lima, R.R., Cannon, J.A., Hsieh, W.W. 2013. Nonlinear regression in environmental sciences by support vector machines combined with evolutionary strategy. Computers & Geosciences 50, 136-144.

Lumbres, R.C., Lee, Y. 2014. Aboveground biomass mapping of La Trinidad forests in Benguet, Philippines, using Landsat Thematic Mapper data and k-nearest neighbor method, Forest Science and Technology. 10:104-111

McNab, W.H. 1989. Terrain shape index: quantifying effect of minor landforms on tree height. Forest Science 35, 91–104.

Molinaro, A. M., Simon, R., Pfeer, R. M., 2005. Prediction error estimation: a comparison of resampling methods. Bioinformatics, Vol. 21, No. 15, pp. 3301-3307.

Moore, I.D., Nieber, J.L. 1989.Landscape assessment of soil erosion and non-point source pollution. J. Minn. Acad. Sci.55, 18–24.

Morel, A.C., Fisher, J.B., Malhi, Y. 2012. Evaluating the potential to monitor aboveground biomass in forest and oil palm in Sabah, Malaysia, for 2000-2008 with Landsat ETM+ and ALOS-PALSAR. International Journal of Remote Sensing 33(11), 3614-3639.

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.

Næsset, E., Gobakken, T., Bollandsås, O.M., Gregoire, T.G., Nelson, R., Ståhl. G., 2013. Comparison of Precision of Biomass Estimates in Regional Field Sample Surveys and Airborne LiDAR-assisted Surveys in Hedmark County, Norway. Remote Sensing of Environment, Vol. 130, pp. 108–120.

O´Loughlin, E. 1986. Predictions of surface saturation zones in natural catchments by topographic analysis. Water Resources Research 22, 794-804.

Porter, F.T., Chen, C., Long, J.A., Lawrance, L.R., 2014. Estimating biomass on crop pastureland: A comparison of remote sensing techniques. Biomass and Bioenergy 66, 268-274.

Price, A., Bauer, B. 1984. Small-scale heterogeneity and soil-moisture variability in the unsaturated zone. Journal of Hydrology. 70, 277-293.

Qiu, Y., Fu, B., Wang, J., Chen, L. 2001. Soil moisture variation in relation to topography and land use in a hillslope catchment of the Loess Plateau, China. Journal of Hydrology 240, 243–263.

Richter R., Schläpfer D. 2011. Atmospheric/topographic correction for satellite imagery. ATCOR 2/3 User Guide. Version 8.0.2.

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.

Roberts, D.W., Cooper, S.V. 1989. Concepts and techniques of vegetation mapping. In: Ferguson, D., Morgan, P., Johnson, F.D. (eds). Land classifications based on vegetation: applications for resource management. Gen. Tech. Rep. INT-257. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station: 90-96.

Sánchez, O., Vega, E., Peters, E., Monrroy, V.O. 2003. Conservación de ecosistemas templados de montaña en México. Instituto Nacional de Ecología (INE–SEMARNAT), México, D.F.

Schume, H., José, G., Katzensteiner, K. 2003. Spatial-temporal analysis of the soil water content in a mixed Norway spruce (*Picea abies (L.) Karst*) – European beech (*Fagus sylvatica L.*) stand. Geoderma 112, 273-287.

SEMARNAT. 2011. Anuario Estadístico de la Producción Forestal 2011.

Seyfried, M. 1998. Spatial variability constraints to modelin soil water at different scales, Geoderma 85, 231-254.

Tucker, J.C. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment. 8:127-150

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. Versión 1.0. Woods Hole Research Center, Falmouth Massachusetts, USA.

Wilson, J.P., Gallant, J.C. Digital terrain analysis. 2000. In *Terrain Analysis: Principles and Applications*, Wilson, J.P., Gallant, J.C., Eds., John Wiley & Sons, Inc.: New York, NY, USA, 2000, pp. 1–28.

Zianis, D., Mencuccini, M. 2004. On simplifying allometric analyses of forest biomass. Forest Ecology and Management 187(2), 311–332.

CAPÍTULO 2. 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]

Abstract

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 (Mg ha^{-1}) 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 modelling 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 variability 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.

[1]Doctorado en Ciencias Agropecuarias y Forestales, DICAF. Universidad Juárez del Estado de Durango.
[2]Instituto de Silvicultura e Industria de la Madera. Universidad Juárez del Estado de Durango.
[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.
[5]División de Estudios de Posgrado e Investigación. Instituto Tecnológico de El Salto.
[6]Departamento de Ingeniería Agroforestal. Universidad de Santiago de Compostela. Escuela Politécnica Superior, Lugo, España.
[*]Author for correspondence: calopez@ujed.mx

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

Quantification of forest biomass and carbon sequestration is an important 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 modelled, 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, transferrable 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 modelling 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 (e.g. Muukkonen & Heiskanen 2005, Fuchs *et al.,* 2009, Hernández-Stefanoni *et al.,* 2011, Aguirre-Salado *et al.,* 2014).

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 & Schläpfer, 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 indices (Gilabert *et al.,* 1997) and indices derived by 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 (e.g. Lu, 2005, Kayitakire *et al.,* 2006, Díaz-Varela *et al.,* 2011) as well as to describe hydrological, geomorphological and ecological processes (e.g. Moore & Nieber, 1989, Wilson & Gallant, 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 nonparametric 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 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 modelling 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).
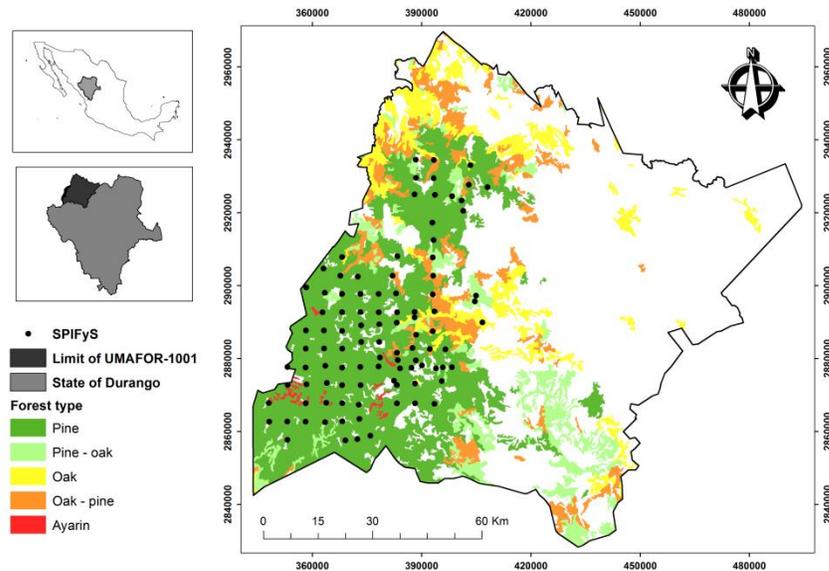
Figure 1. Geographical location of the study area and sample plots used in the study.

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 between three and five kilometres, depending on the orography of the study area. In each plot (50 x 50 metres), 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) aboveground 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.

Table 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 ($m^2 ha^{-1}$)	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

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, at 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-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 corrected radiometrically, atmospherically and topographically using the ATCOR3[®] module, regarded as particularly suitable for mountainous zones, (Geosystems 2013) and implemented with ERDAS[®] IMAGINE[®] 2013 software (ERDAS Inc. 2013). After the correction, the original image digital levels (DL) were converted to ground reflectance values for each band. A number of vegetation indices 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 indices 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 vegetation canopy biomass. Hence, the Normalized Difference Vegetation Index (Rouse *et al.,* 1974) and Soil Adjusted Soil Vegetation Index (Huete 1988) were included as indices 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.

Table 2. Vegetation indices analyzed in the present study.

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

where:
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
$\int_{0.3\mu m}^{2.5\mu m}$: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

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 vapour 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 with a spatial resolution of 15 m (INEGI, 2014). Prior to these calculations, a low pass filter with a 5x5 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 ERDAS® IMAGINE® 2013 software (ERDAS Inc. 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 3x3, 5x5 and 7x7 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 PCI Geomatica2013® software (PCI Geomatics Inc. 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 modelled (McNab, 1989, Roberts &

Cooper, 1989). Therefore, first and second order terrain parameters (Tab. 3) were derived from the 5x5 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.

Table 3. Additional variables for biomass modelling.

Group variable	Variable	Formula	Reference
Texture (NDVI)	Homogeneity (HO)	$HO = \sum_{i,j=0}^{N-1} i \frac{P_{i,j}}{1+(i-j)^2}$	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 (Sdt)	$Sdt = \sqrt{VA}$	
	Entropy (EN)	$EN = \sum_{i,j=0}^{N-1} i\, P_{i,j}\,[-lni - P_{i,j}]$	
	Angular Second Moment (ASM)	$ASM = \sum_{i,j=0}^{N-1} i\, P^2{}_{i,j}$	
	Correlation (CR)	$CR = \sum_{i,j=0}^{N-1} i\, P_{i,j}\left[\frac{(i-ME)(j-ME)}{\sqrt{VAi\,VAj}}\right]$	
Terrain (DEM)	Elevation	Digital Elevation Model	
	Slope (β)	$\beta = \arctan\left[(G^2+H^2)^{\frac{1}{2}}\right]$	Roberts and Cooper (1989)
	Transformed Aspect (Trasp)	$Trasp = \dfrac{1-\cos((\pi/180)(\alpha-30))}{2}$	
	Terrain Shape Índex (TSI)	$TSI = \dfrac{\bar{Z}}{R}$	McNab (1989)
	Wetness Index (WI)	$W = \ln(As/\tan\beta)$	Moore and Nieber (1989)
	Profile curvature (Ø)	$\varnothing = -2\dfrac{DG^2+EH^2+FGH}{G^2+H^2}$	Wilson and Gallant (2000)
	Plan curvature (ω)	$\omega = 2\dfrac{DH^2+EG^2+FGH}{G^2+H^2}$	
	Curvature (x)	$x = \omega - \varnothing$	

where:
N: Is the number of grey levels
P: Is the normalized symmetric GLCM of dimension N x N.
V: Is the vector difference normalized grey level of dimension N.
P (i, j): Is the matrix of co-occurrence normalized so that $\sum_{i,j=0}^{N-1} \sum_{i,j=0}^{N-1} P\ (i-j)$.
V (k): Is the normalized grey level difference vector $\sum_{i,j=0}^{N-1} \sum_{i,j=0}^{N-1} P\ (i-j)|\,i\text{-}j\,|=k$
Z: Average elevation.
R: Point radio altitude units.
As: Drainage area specified.
tanβ: Local slope angle.
VA: Variance.
ME: Mean.
D, F, G and H were derived according to equation of Zevenbergen and Thorne (1987).

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 R statistical software (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 indices (6 indices), the texture variables derived from the Normalized Difference Vegetation Index (NDVI) (24 variables), and the terrain variables derived from the DEM (9 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 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:

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

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 $\frac{1}{N_t}\sum_{i=1}^{N_t}(y_i - \bar{y}_t)^2$ 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 analysing 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 evaluation of model performance: the coefficient of determination (R^2) and the RMSE. The expressions of these statistics are summarized as follows:

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

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 multicollinearity. This refers to the existence of high correlations between certain independent variables if they represent or measure 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, p. 168), an assessment of the validity of the models with an independent data set 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 Inc. 1999-2012) 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 variability 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 5x5 window and *wetness index*.

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

Parameter	Estimate	Standard error	OLS RMSE (Mg ha^{-1})	R^2	Cross-validation 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	0.5833	30.31	0.5133
LAI	0.0166	0.0037	*31.32%		*34.04%	
Modified SAVI	-2215.8205	717.3984				
NDVI	2072.8881	628.6811				
Wetness index	3.4027	1.5666				
Contrast 7x7	0.2073	0.0926				
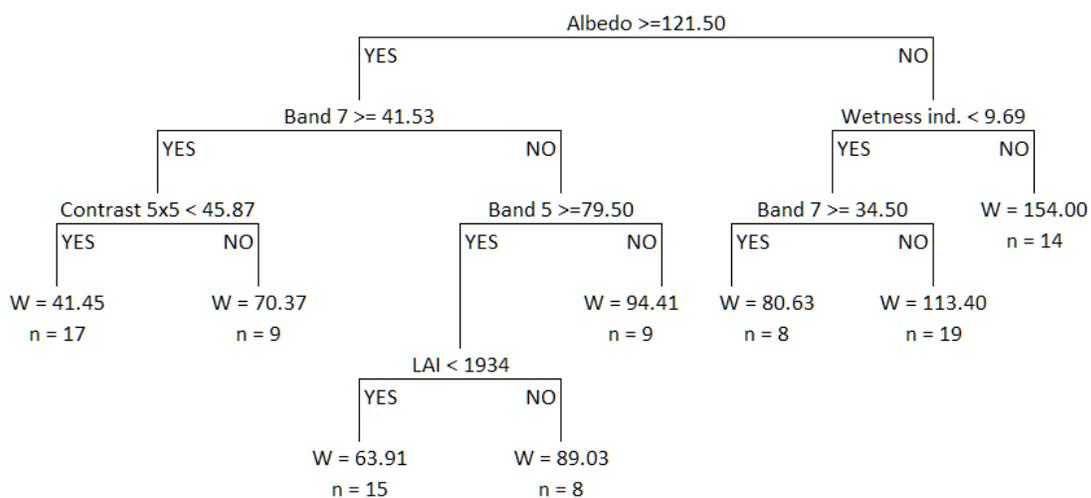
Figure 2. Classification tree obtained by the regression tree method *n* is the number of sample plots in each node and *W* is the biomass value for each node (Mg ha^{-1}).

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 removing branches or terminal nodes that contribute little to explaining the response variable, to yield an appropriately-sized 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*).
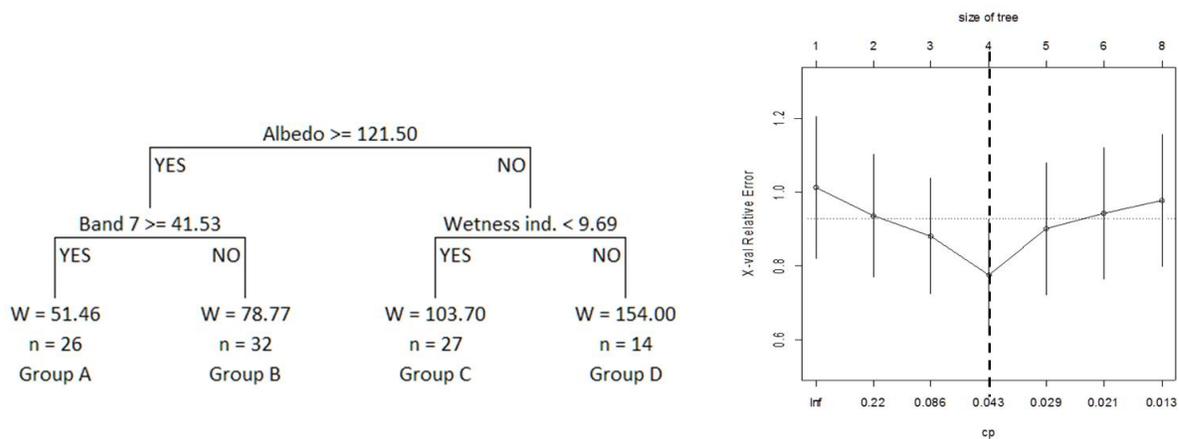
Figure 3. Classification tree obtained by pruning the regression tree (left) and plot of the relationships between the cost-complexity parameter (CP), the cross validation error (x-val Relative Error) and tree size (number of nodes). The dashed vertical line represents the maximum number of nodes (corresponding cost-complexity parameter) for which the cross validation error is greater than the standard error (right).

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 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).

Table 5. Model obtained for the final nodes of the regression tree by OLS and stepwise selection of independent variables ([*]RMSE is 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 (Mg ha^{-1})	R^2	[**]RMSE reduction (%)
A	Abs. Shortw. solar rad.	0.1004	0.0523			
	Entropy 3x3	-143.8093	32.8994	19.85	0.2605	14.73%
	Correlation 3x3	-35.0613	11.0553	[*]38.57%		
	Dissimilarity 5x5	12.3776	2.9020			
	Mean 7x7	1.6886	0.4856			
B	Band 1	4.5965	0.7375	16.87	0.2591	18.97%
	Band 5	-2.1912	0.5013	[*]21.42%		
C	Intercept	360.9605	102.8189	26.71	0.1382	1.05%
	Band 1	-4.7552	1.9528	[*]25.76%		
	Correlation 3x3	-137.3135	41.1643			
D	Flow solar rad.	-0.2223	0.0667	31.38	0.6149	46.14%
	NDVI	745.2892	112.9641	[*]20.38%		
	Stand. Dev. 7x7	-7.0717	2.0254			

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 variability in stand biomass, with a RMSE value of 22.59 Mg ha^{-1} (25.37% of the mean stand biomass).
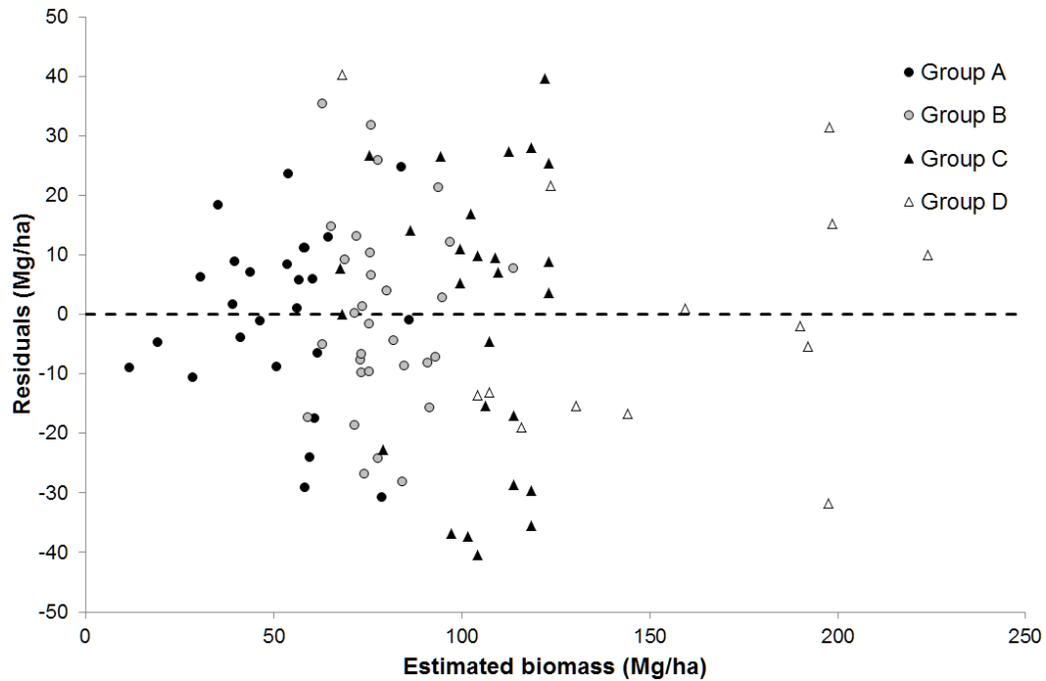
Figure 4. Plot of residual values against estimated biomass for groups obtained from the classification tree.

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 blue and yellow pixels reflect biomass contents of between 52 and 168 Mg ha[-1], whereas the green pixels (light and dark) represent the highest biomass values found in temperate forest (mainly pine and pine-oak cover) in the study area, in accordance with INEGI's Land Use and Vegetation chart, series V (INEGI 2012).
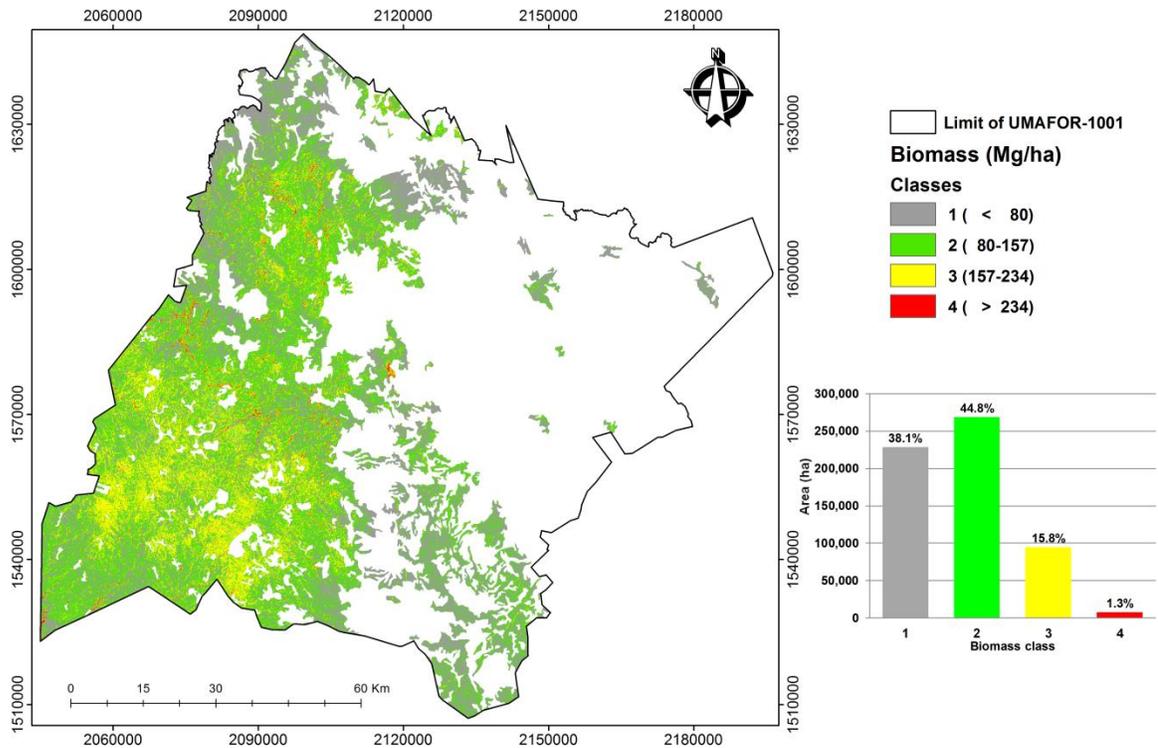
Figure 5. Biomass distribution in the study area.

Discussion

The results of the present research showed that integration of spectral information, texture variables derived from the NDVI and terrain indices (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 (e.g. 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 of 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 index. 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 *Pinus 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 (e.g. Moore *et al.,* 1993, Zinko *et al.,* 2005) and of potential areas for establishing forest plantations (e.g. 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 *et al.,* 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 indices, 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 sensing 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 3x3, Correlation 3x3, Dissimilarity 5x5, Mean 7x7,, Stand. Dev. 7x7) were included in the mixed model. This indicates the importance of the spatial arrangement of spectral values at different spatial scales (from a 3x3 kernel corresponding with an area of 0.81 ha to a 7x7 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 indices 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 500x500 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 indices 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

References

Aguirre-Salado, C.A,, Aguirre-Salado, E.J., Treviño-Garza, O.A., Aguirre-Calderón, J., Jiménez-Pérez, M.A., González-Tagle, J.R., 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.

Asrar G (1989). Theory and applications of optical remote sensing, J. Wiley & Sons, New York.

Asrar G, Fuchs M, Kanemasu ET, Hatfield, J.L. 1984. Estimating absorbed photosynthetically active radiation and leaf area index from spectral reflectance in wheat. Agronomy Journal 76(2): 300-306.

Balthazar, V. Veerle, V. Eric, F.L. 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.

Bannari, A., Morin, D., Bonn, F., Huete, A.R. 1995. A review of vegetation indices. Remote Sensing Reviews.13: 95–120.

Baret, F. and Guyot, G. 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sensing of Environment. 35: 161-173.

Belsey, D.A. 1991. Conditioning diagnostics, collinearity and weak data in regression. John Wiley & Sons Inc., New York.

Botero, V.J.S. and Restrepo, M.A. 2010. Análisis de textura en panes usando la matriz de coocurrencia. Revista Politécnica 6(10): 74-80.

Breiman. L., Friedman, J.H., Olshen, R.A., Stone, C.J. 1984. Classification and Regression Trees. Chapman & Hall, New York.

Brown, S., Lugo, A.E. 1984. Biomass of tropical forests: a new estimate based on forest volumes. Science 223(4642): 1290-1293.

Brutsaert, W, 1975. On a derivable formula for long-wave radiation from clear skies, Water Resources Research 11: 742-744.

Challenger, A. 1998. Utilización y conservación de los ecosistemas terrestres de México: pasado, presente y futuro. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad/Instituto de Biología, UNAM/Agrupación Sierra Madre, México, D.F.

Chen, D., Stow, D.A., 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.

Corral-Rivas, J. Vargas, B., Wehenkel, C., Aguirre, O., Álvarez, J.G., Rojo, A. 2009. Guía para el Establecimiento de Sitios de Inventario Periódico Forestal y de Suelos del Estado de Durango. Facultad de Ciencias Forestales. Universidad Juárez del Estado de Durango.

Cutler, M.E.J., Boyd, D.S., Foody, G.M., 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.

Díaz-Varela, R.A., 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.

ERDAS Inc 2013. Erdas Imagine 2013. http://www.hexagongeospatial.com/products/ERDAS-IMAGINE/details.aspx.

ESRI Inc. 1999-2012. ArcGIS for Desktop 10.

Foody, G.M., Boyd, D.S., Cutler, M.E.J. 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. Remote Sensing of Environment 85(4): 463–474.

Forzieri, G., Feyen, L., Cescatti, A., Vivoni, E.R. 2014. Spatial and temporal variations in ecosystem response to monsoon precipitation variability in southwestern North America, J. Geophys. Res. Biogeosci., 119.

Franklin, S.E., Maudie, A.J., Lavigne, M.B. 2001. Using Spatial Co-Occurrence Texture to Increase Forest Structure and Species Composition Classification Accuracy. Photogrammetric Engineering and Remote Sensing 67: 849–856.

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.

García, M.A., Pérez, C.F., De la Riva, J., Fernández, J., Pascual, P.E., Herranz, A. 2005. Estimación de la biomasa residual forestal mediante técnicas de teledetección y SIG en masas puras de *Pinus halepensis* y *P. sylvestris*. In: IV Congreso Forestal Español. Sociedad Española de Ciencias Forestales. 4CFE05-342-T1, 308.

Geosystems, 2013. Haze Reduction, Atmospheric and Topographic Correction. User Manual ATCOR2 and ATCOR3. 2013. Geosystems GmbH. http://www.geosystems.de/atcor/

Gilabert, M.A., Piqueros, J.G., Haro, J.G. 1997. Acerca de los índices de vegetación. Revista de Teledetección 8: 1-10.

Glenn, E.P., Huete, A.R., Nagler, P.L., Nelson, S.G. 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.

González-Alonso, F., Merino-De-Miguel, S., Roldán-Zamarrón, A., García-Gigorro, S., Cuevas, J.M. 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.

Haralick, R.M., Shanmugam, K., Dinstein, I. 1973. Textural features for image classification. IEEE Transactions of Systems, Man, and Cybernetics 6: 610-621.

Hernández-Stefanoni, J.L., Gallardo-Cruz, J.A., Meave, J.A., Dupuy, J.M. 2011. Combining geostatistical models and remotely sensed data to improve tropical plant richness mapping. Ecological Indicators 11: 1046–1056.

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.

Houghton, R.A., Butman, D., Bunn, A.G., Krankina, O.N., Schlesinger, P., Stone, T.A. 2007. Mapping Russian forest biomass with data from satellites and forest inventories. Environ. Res. Lett. 2: 1-7

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.

Huete, A.R. 1988. A Soil-Adjusted Vegetation Index (SAVI). Remote Sensing of Environment 25: 295-309.

INEGI - Instituto Nacional de Estadística Geográfica e Informática. 2012. Uso del Suelo y Vegetación Escala 1: 250 000 Serie V, Información vectorial. México.

INEGI - Instituto Nacional de Estadística Geográfica e Informática. 2014. http://www.inegi.org.mx/geo/contenidos/datosrelieve/continental/descarga.aspx. 5-06-2014.

IPCC. 2003. Good practice guidance for land use, land-use change and forestry. Hayama, Japan: IPCC National Greenhouse Gas Inventories Programme.

Kayitakire, F., Hamel, C., Defourny, P. 2006. Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery. Remote Sens. Environ. 102: 390–401.

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.

Ketterings, Q.M., Coe, R., Van, M.N., Ambagau, Y., Palm, C.A. 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: 199–209.

Keys, R.G. 1981. Cubic convolution interpolation for digital image processing. IEEE Transactions on Acoustics, Speech, and Signal Processing 29: 1153–1160.

Kozak, A., Kozak, R.A. 2003. Does cross validation provide additional information in the evaluation of regression models? Canadian Journal of Forest Research 33: 976-987.

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.

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.

Liang, S. 2000. Narrowband to broadband conversions of land surface albedo I algorithms. Remote Sensing of Environment 76: 213-238.

Liang, S. 2007. Recent developments in estimating land surface biogeophysical variables from optical remote sensing. Progress in Physical Geography 31: 501-516.

Lu, D. 2005. Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon Basin. International Journal of Remote Sensing. 26: 2509–2525.

McNab, W.H. 1989. Terrain shape index: quantifying effect of minor landforms on tree height. Forest Science 35: 91–104.

McNab, H.W. 1993. A topographic index to quantify the effect of mesoscale landform on site productivity. Canadian Journal of Forest Research 23: 1100-1107.

Méndez-Barroso, L. A., Vivoni, E.R., Watts, C.J., Rodríguez, J.C. 2009. Seasonal and interannual relation between precipitation, surface soil moisture and vegetation dynamics in the North American monsoon region. J. Hydrol. 377, 59–70.

Meyer, P., Itten, K.I., 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.

Moore, I.D., Nieber, J.L. 1989. Landscape assessment of soil erosion and non-point source pollution. Journal of the Minnesota Academy of Science. 55(1): 18-24.

Moore, I.D., Norton, T.W., Williams, J.E. 1993. Modelling environmental heterogeneity in forested landscapes. Journal of Hydrology. 150: 717–747.

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.

Myers, R.H. 1990. Classical and modern regression with applications. 2nd edition. Duxbury Press, Belmont, California.

NASA. 2011. Landsat 7 Science Data Users Handbook. NASA, URL: http://landsathandbook.gsfc.nasa.gov/pdfs/Landsat7_Handbook.pdf . pp. 30-31.

Neter, J., Wasserman, W., Kutner, M.H. 1990. Applied linear statistical models: regression, analysis of variance and experimental designs. 3rd ed. Irwin, Boston, Mass.

Nichol, J.E. and Sarker, M.L.R. 2011. Improved Biomass Estimation Using the Texture Parameters of Two High-Resolution Optical Sensors. IEEE Transactions on Geoscience and Remote Sensing. 49: 930–948.

PCI Geomatics Inc. 2013. PCI Geomatics 2013. http://www.pcigeomatics.com/

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H. 1994. Modified Soil Adjusted Vegetation Index (MSAVI). Remote Sensing of Environment. 48:119-126.

R Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from 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.

Richter, R. and Schläpfer, D. 2011. Atmospheric/topographic correction for satellite imagery. ATCOR 2/3 User Guide. Version 8.0.2.

Roberts, D.W., Cooper, S.V. 1989. Concepts and techniques of vegetation mapping. In: In: Ferguson, D., Morgan, P. and Johnson, F.D. (eds). Land classifications based on vegetation: applications for resource management. Gen. Tech. Rep. INT-257. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station: 90-96.

Rouse, J.W., Haas, R.H., Schiell, J.A., Deferino, D.W., Harlan, J.C. 1974. Monitoring the vernal advancement of retrogradation of natural vegetation. NASA/OSFC. Type III. Final Report.Oreenbello MD.

Ryu, S., Chen, J., Crow, T.R., Saunders, S.C. 2004. Available fuel dynamics in nine contrasting forest ecosystems in North America. Environmental Management. 33(1): 87–107.

Salinas-Zavala, C.A., Douglas, A.V., Diaz, H.F. 2002. Interannual variability of NDVI in northwest Mexico: Associated climatic mechanisms and ecological implications. Remote Sens. Environ. 82: 417–430.

Sánchez, O., Vega, E., Peters, E., Monrroy–Vilchis, O. 2003. Conservación de ecosistemas templados de montaña en México. Instituto Nacional de Ecología (INE–SEMARNAT), México, D.F.

SAS Institute Inc. 2007. SAS OnlineDoc, Versión 9.2, Cary, Carolina del Norte: SAS Institute Inc.

Sun, G., Ranson, J.K. 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, K.J., Guo, C., Zhang, Z., Montesano, P., Kimes, D. 2011. Forest biomass mapping from lidar and radar synergies. Remote Sensing of Environment 115: 2906–2916

Therneau, T.M., Atkinson, B. 2012. Package rpart. Version 3.1-52. http://cran.r-project.org/web/packages/rpart/index.html Accessed 12 March 2014.

Vargas-Larreta, B. 2013. Estimación del potencial de los bosques de Durango para la mitigación del cambio climático. Modelización de la biomasa forestal. Proyecto FOMIX-DGO-2011-C01-165681. Comisión Nacional de Ciencia y Tecnología (CONACYT). 53 pp.

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. Versión 1.0. Woods Hole Research Center, Falmouth Massachusetts, USA.

Wilson, J.P., Gallant, J.C. 2000. Digital terrain analysis. In: Wilson, J.P. and Gallant, J.C. (eds.). Terrain Analysis: Principles and Applications. John Wiley&Sons, Inc.,New York, pp. 1–28.

Zevenbergen, L.W., Thorne, C.R. 1987. Quantitative analysis of land surface topography. Earth Surface Processes and Landforms. 12: 47-56.

Zhao, M., Running, S.W. 2010. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. Science. 329: 940–943.

Zianis, D., Mencuccini, M. 2004. On simplifying allometric analyses of forest biomass. Forest Ecology and Management. 187(2): 311–332.

Zinko, U., Seibert, J., Dynesius, M., Nilsson, C. 2005. Plant species numbers predicted by a topography based groundwaterflow index. Ecosystems. 8: 430-441.

CAPÍTULO 3. GEOSPATIAL ESTIMATION OF ABOVE GROUND FOREST BIOMASS IN THE SIERRA MADRE OCCIDENTAL IN THE STATE OF DURANGO, MEXICO

Pablito M. López-Serrano[1], Carlos A. López-Sánchez[2]*, Raúl Solís-Moreno[3] and José J. Corral-Rivas[2]

Abstract

Combined use of new geospatial techniques and non-parametric multivariate statistical methods enables monitoring and quantification of the biomass of large areas of forest ecosystems with acceptable reliability. The main objective of this study was to estimate above-ground forest biomass (AGB) in the Sierra Madre Occidental (SMO) mountain range in the state of Durango, Mexico, by analysis of medium-resolution spectral data and field data collected from a network of 201 permanent monitoring sites (SPIFyS) installed by systematic sampling in the area in 2011. The digital levels of the images were converted to apparent reflectance (ToA) and surface reflectance (SR). Was used the algorithm M5P that constructs tree-based piecewise linear models. The fitted model explained 69% of the variance in the observed AGB (RMSE=42.17 Mg ha^{-1}). The variables that best discriminated the AGB, in order of decreasing importance, were the spectral bands in the red and near-infrared (NDVI), the mid-infrared and the blue regions. The results demonstrate the potential usefulness of the M5P method for estimating AGB based in the surface reflectance values (SR).

Keywords: M5P, Remote Sensing, Landsat, Sierra Madre Occidental.

[1]Doctorado Institucional en Ciencias Agropecuarias y Forestales, Universidad Juárez del Estado de Durango,
[2]Instituto de Silvicultura e Industria de la Madera, Universidad Juárez del Estado de Durango
[3] Facultad de Ciencias Forestales, Universidad Juárez del Estado de Durango
*Author to whom correspondence should be addressed, E-Mail: calopez@ujed.mx.

Introduction

The Sierra Madre Occidental (SMO) mountain range is of great ecological interest because of its environmental heterogeneity, which is attributed to the broad physiographical and climatic diversity in the area (Gonzalez *et al.,* 2012). Moreover, the SMO is home to pine and oak species that are economically important in ecosystems in Mexico and other parts of the world (Sànchez *et al.,* 2003). The SMO crosses several states in western Mexico, including the state of Durango (the SMO occupies 71.5% of the surface area of the state). The state of Durango generates between 25% and 30% of the national timber production, producing a total of 1.5 million·m^3 of roundwood per year, and boasts forest reserves that are important sources of environmental services (SEMARNAT, 2011). Studies that attempt to estimate forest biomass in this type of ecosystem are expensive due to its large coverage and difficult access for direct estimation of biomass. Thus, the emergence of geospatial techniques is becoming increasingly relevant for estimating and monitoring forest biomass in short periods of time because of its low cost and acceptable accuracy (Hall *et al.,* 2006, Fuchs *et al.,* 2009; Verbesselt *et al.,* 2010). Because of the macrospatial scale and high heterogeneity of these ecosystems, the quantitative data obtained often do not comply with the underlying assumptions of simple statistical analysis (homogeneity and normality of distribution), so other techniques such as logistic regression and non-parametric classification methods are often applied (Gibbons and Chakraborti, 2003; Barajas *et al.,* 2007; Karjalainen *et al*., 2012). The M5 model tree (M5P) technique is a reconstruction of M5 algorithm for inducing trees of regression models (Quinlan, 1992). M5P is used for numeric prediction and at each leaf it stores a linear regression model that predicts the class value of instances that reach the leaf. To determine which attribute is the best to split the portion of the training data that reaches a particular node, the splitting criterion is used. The standard deviation of the training class is treated as a measure of the error at that node and each attribute at that node is tested by calculating the expected reduction in error. The attribute that is chosen for splitting maximizes the expected error reduction at that node. The main objective of the present study was to estimate the aboveground forest biomass (AGB) in the SMO in the state of Durango, Mexico, using the M5P technique and the analysis of medium-resolution satellite-based multi-spectral data, and field data collected from a network of 201 permanent forest growth and soil research sites (SPIFyS).

Material and Methods

Study Area

The study area is a mountainous zone in the state of Durango (Mexico) that forms part of the Sierra Madre Occidental (Figure 1).
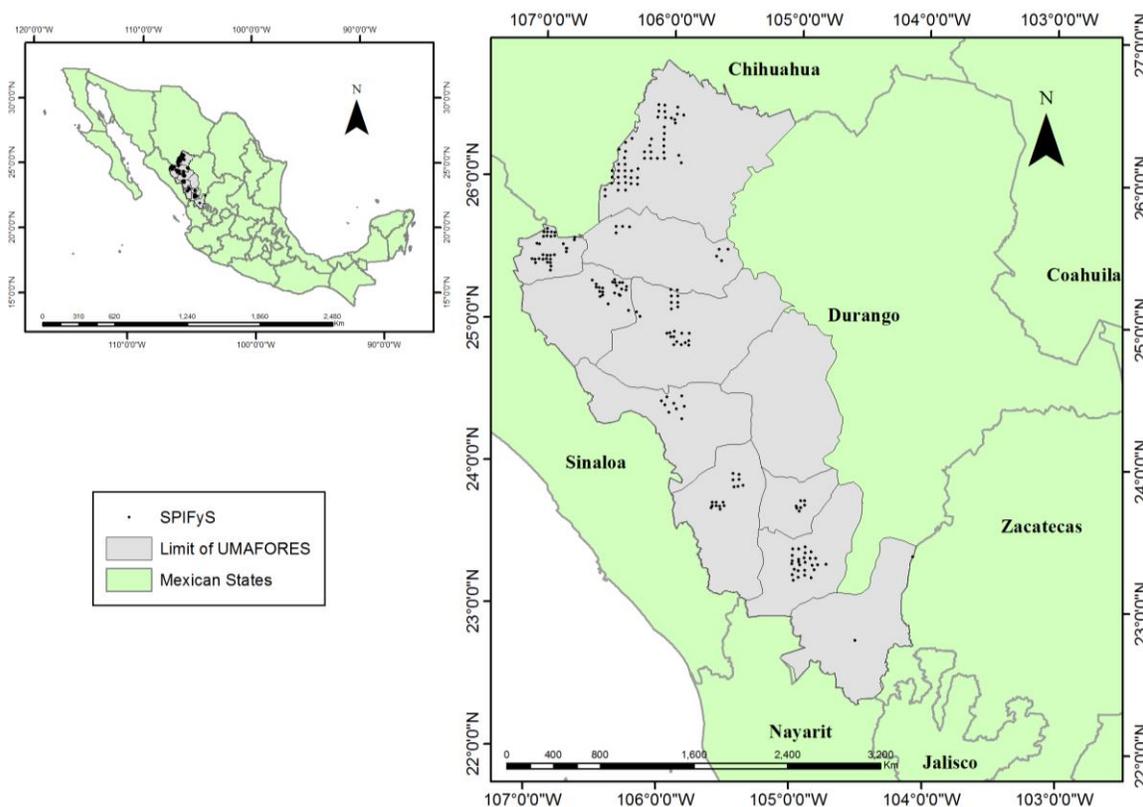
Figure 1. Location of the study area.

The area occupied by the state of Durango represents 6.3% of the land in Mexico. The total area covered by the state is 12.3 million ha, of which 9.1 million ha (74.35% of the land in the state) is forestland managed by 11 Regional Forest Management Units (UMAFORES). A large part of the forestland (4.9 million ha) is occupied by temperate forest and is subjected to precipitation levels of between 800 and 1200 mm per year, with frost occurring in winter as a result of the combination of low temperatures and humid winds from the Pacific Ocean, a smaller area of the land (0.5 million), affected by warmer climate, is occupied by forest classified as rainforest (SRNyMA-CONAFOR, 2007). The mean elevation in this zone is 2650 m above sea level. These forests have rich biodiversity and include at least 27 coniferous tree

species (of which 20 are *Pinus* species) and 43 species of *Quercus*, the predominant forest stands comprise pines and oaks, often mixed with *Arbutus* and *Juniperus*, among other tree species (Zhao *et al.,* 2014). These unique forests are irregular and have been subject to selective harvesting for almost a century to provide a mix of services to local communities. This irregularity refers to the spatial arrangement of trees (vertical and horizontal irregularity) and the variation in the age structure of trees and stands. This structure is the result of the management history, which has depended on land ownership, as well as the economic and social changes that have taken place in the state, and also natural conditions (Wehenkel *et al.,* 2011).

Field Data

The dasometric data were obtained from 201 permanent forest growth and soil monitoring plots (SPIFyS) in the SMO in the state of Durango. The plots were installed during the winter of 2011 using the protocol developed by (Corral-Rivas *et al*., 2009). The data of these permanent sample plots are used to monitor the growth and yield of Durango's forests. The plots cover the main forest types and the current diameter distributions of commercial forests in Durango. The plots are 50 × 50 m in size (distance was corrected by the slope) and are distributed by systematic sampling (with some exceptions), with a variable grid ranging from 3 to 5 kilometers, depending on the size of the "*Ejidos*". *Ejidos* are communal groups that live in rural areas and whose lands are managed with some level of governmental control. The sampling plots are intended to be re-measured at five-year intervals. Among other variables, tag number, species code, breast height diameter (measured in cm at 1.3 m above ground level), total tree height (m), height to the live crown (m), azimuth (°) and radius (m) from the center of the plot of all trees equal or larger than 7.5 centimeters (cm) in diameter were recorded. The database used here includes measurement data from 31,979 trees.

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

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

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

Variable	Mean	Standard Deviation	Minimum Value	Maximum Value
Number of stems per ha	645	271.84	224	2264
Stand basal area ($m^2 \cdot ha^{-1}$)	23.44	8.06	8.21	54.83
Dominant height (m)	17.47	5.08	6.86	30.60
Stand biomass ($Mg \cdot ha^{-1}$)	141.64	75.01	27.73	469.42

Tree Abundance by Species Group (ASG)

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

Source of Spectral Data

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

The satellite images were digitally pre-processed by radiometric correction techniques, according to the procedures suggested by (Greenle, 1993; Chuvieco, 2010). The images are produced by USGS with a rectification using a cubic convolution geometric correction for discrete data (level L1T), with a root mean square error (RMSE) of less than 1 pixel, thus making them suitable for digital image processing (Keys, 1981). The digital levels (DLs) were converted to radiance values to generate images that were calibrated with the minimal radiance (Lmin) and

maximal radiance (Lmax) values for each band of the sensor (Eastman, 2012). The radiance was subsequently converted to apparent reflectance (Top of Atmosphere (ToA)) with the aim of converting the original values of each image into standard physical variables that are comparable over time for the same sensor (Greenle, 1993). This process was carried out with IDRISI$^®$ Selva software (Eastman, 2012) and the ATMOS algorithm, which fits the radiometric effect on considering the solar elevation, yielding an image with reflectance values (0–1).

Furthermore, the same images were downloaded from the National Landsat Archive Processing System (NLAPS), corresponding to the product Landsat 4–5 Thematic Mapper level 1 of reflectance on surfaces (SR), radiometrically and atmospherically corrected, and processed through the Standard Landsat Product Generation System (LPGS) using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (USGS, 2011).

Bands 1, 2, 3, 4, 5 and 7 (Band 1 to Band 7) of Landsat-5 TM were used, Band 6 was not used, because of its thermal characteristics (USGS, 2011). The Normalized Difference Vegetation Index (NDVI) was also calculated, with the aim of compensating the factors that influence the images in relation to biomass estimation, such as the illumination conditions, the slope and the orientation of the surface. The use of NDVI calculated for ToA and SR, as a predictor variable to model AGB has been successfully reported in previous studies (Gasparri *et al*., 2010; Liang *et al*., 2012; Zhu and Liu, 20015).

$$NDVI = \frac{(NIR - R)}{(NIR + R)} \tag{1}$$

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

Integration of Data Files

Once the images were obtained by the previously mentioned processes (ToA and SR spectral bands), a mosaic was constructed with six of the scenes covering the SMO. Posterior geolocation of the SPIFyS in the mosaic enabled extraction of the information at the pixel level by bilinear interpolation. ArcGIS 10$^®$ software (ArGIS, 2012) was used for this extraction. ToA and SR values were integrated in a database together with the extracted total aboveground biomass (Mg·ha^{-1}) as inputs for the model.

Fitted Model

We used a machine learning technique to estimate the AGB at stand level. M5P technique combines a conventional decision tree with the possibility of linear regression functions at the nodes. First, a decision-tree induction algorithm is used to build a tree, but instead of maximizing the information gain at each inner node, a splitting criterion is used that minimizes the intra-subset variation in the class values down each branch. The splitting procedure in M5P stops if the class values of all instances that reach a node vary very slightly, or only a few instances remain. Second, the tree is pruned back from each leaf. When pruning, an inner node is turned into a leaf with a regression plane. Third, to avoid sharp discontinuities between the subtrees, a smoothing procedure is applied that combines the leaf model prediction with each node along the path back to the root, smoothing it at each of these nodes by combining it with the value predicted by the linear model for that node. Techniques devised by (Quinlan, 1992) for their classification and regression trees system are adapted in order to deal with enumerated attributes and missing values. All enumerated attributes are turned into binary variables so that all splits in M5P are binary. As to missing values, M5P uses a technique called "surrogate splitting" that finds another attribute to split on in place of the original one and uses it instead (Breiman *et al.*, 1984; Quinlan, 1992; Wnag and Witten, 1997).

In this study, in a first stage, the six spectral bands of the Landsat 5 TM sensor (1, 2, 3, 4, 5 and 7) and the NDVI were analyzed with the algorithms ToA and SR to estimate AGB. In a second stage, its spectral variables (SR) were evaluated with variables that incorporate aspects of forest structure (ASG). All analyses were performed with M5P technique implemented into the WEKA open source software (Hall, 1999).

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

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}} \tag{3}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2}} \tag{4}$$

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

The selected model was applied for mapping AGB in the SMO area using ArcGIS $10^{®}$ software (ArGIS, 2012)

Results

The decision tree generated by the M5P technique for ToA, SR and SR with ASG variables were implemented in WEKA software, using the pixel level values extracted from the images of the 201 SPIFyS plots is shown in Figure 2.
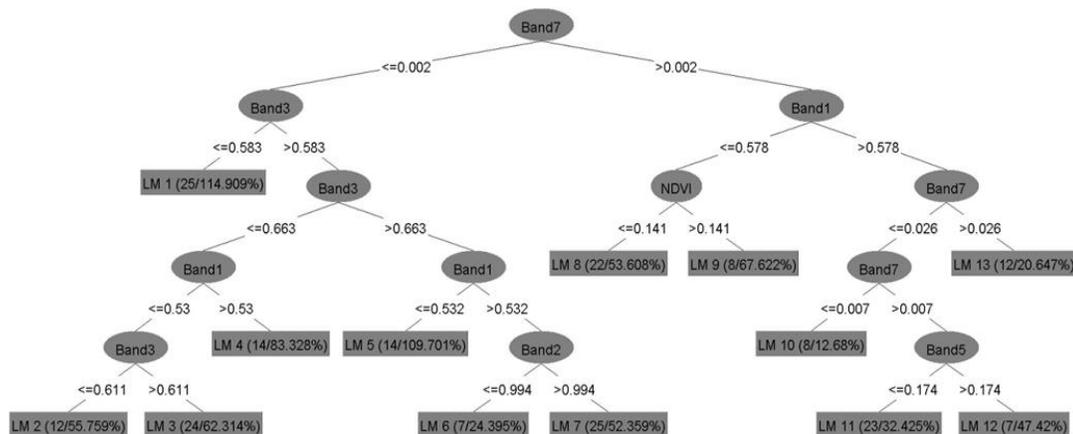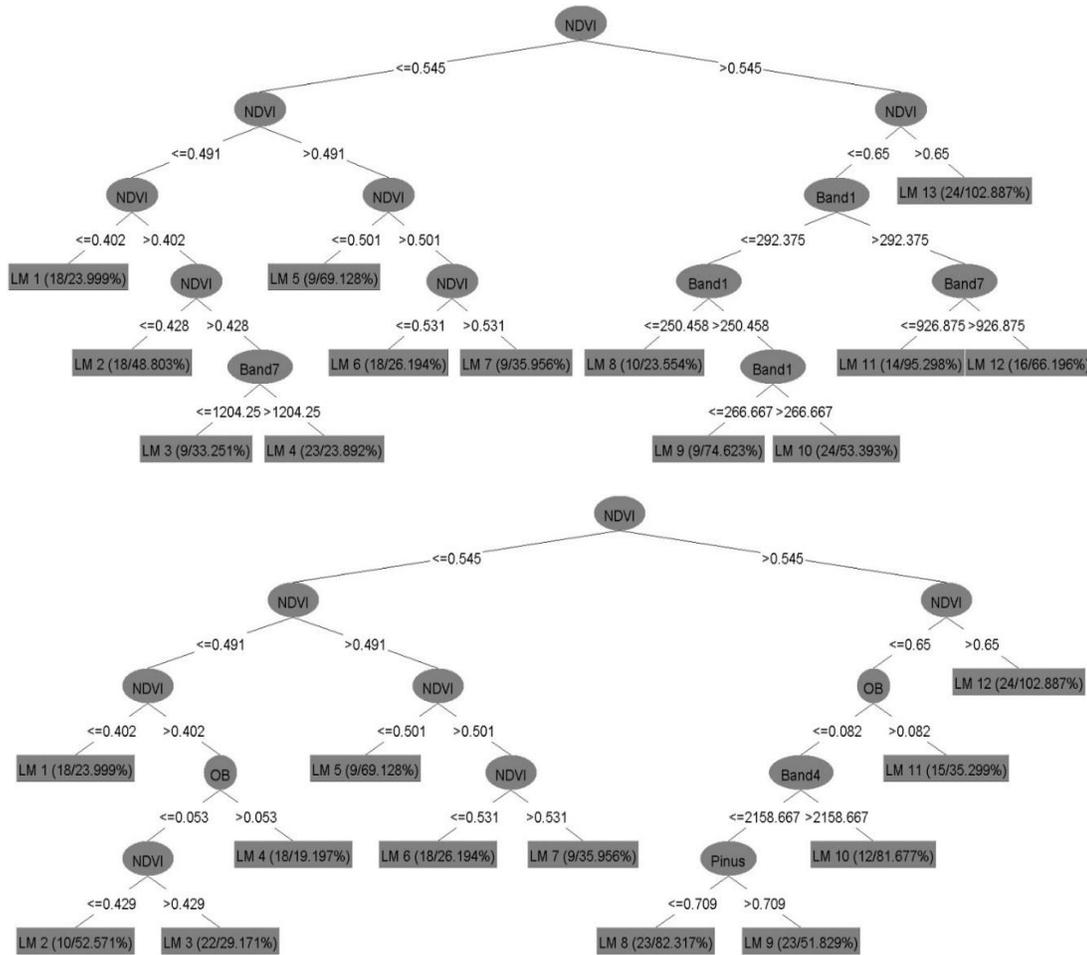
Figure 2. Decision tree obtained using the M5P technique with ToA (upper), SR (middle) and SR with ASG variables (bottom). (Band 1 to 7): of Landsat 5 TM (Thematic Mapper) satellite, (NDVI): normalized difference vegetation index, (OB) tree abundance of other broadleaves species, and (Pinus) tree abundance of pines.

In accordance with the hierarchical structure of the decision trees, the following variables that best discriminated or predicted the AGB, in order of decreasing importance were, for ToA: Band 7, Band 3, Band 1, NDVI and Band 5, for SR: NDVI, Band 1 and Band 7, and for SR with ASG: NDVI, OB, Band 4 and tree abundance of pines. Categorization of the trees continued following the path determined by the responses to the questions at the internal nodes, until reaching a terminal node, where the predetermined label will be that assigned to the classification pattern—in this case, the pixel values for AGB estimation. The Table 2 show the goodness-of-fit statistics derived from the M5P technique with ToA values explained 54% (R^2) of the observed variability in the AGB of the 201 research plots, with a RMSE of 50.47

Mg·ha^{-1}. SR model explained 69% (RMSE = 42.17 Mg·ha^{-1}), and when including the ASG variables the explanation of the variance increases to 73% (RMSE = 39.40 Mg·ha^{-1}).

Table 2. Summary of the goodness-of-fit statistics for estimation of the AGB.

Statistics	ToA	SR	SR with ASG
R^2	0.54	0.69	0.73
RMSE	50.47	42.17	39.40
RRSE	67.45	56.36	52.66

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

The spatial distribution of the estimated AGB (Mg·ha^{-1}) in the SMO area obtained by the application of the classification rules included in the regression tree model (M5P) for SR variables is shown in Figure 4. The lighter color pixels represent the lowest amounts of AGB, below 75 Mg·ha^{-1}, whereas the dark green pixels represent the largest amounts of AGB, which consistently correspond to the most dense areas of temperate forest. Calculated mean amount of AGB for the study area was around 106 Mg·ha^{-1}.

The total AGB content estimations from the MP5 technique for SR variables for the analyzed forest management units are shown in Table 3. The highest mean value of AGB was observed in the UMAFOR 1006 (Municipally of San Dimas) with 148.98 Mg·ha^{-1} and a total estimation of 64,033,008.59 Mg. This zone encompasses the largest area of forestland and therefore the largest amount of AGB. On the other hand, the lowest amount of AGB was observed in the UMAFOR 1001 with a mean estimate of 78.66 Mg·ha^{-1}, making it the forest region with the lowest density out of the eleven forest management units considered in this study.
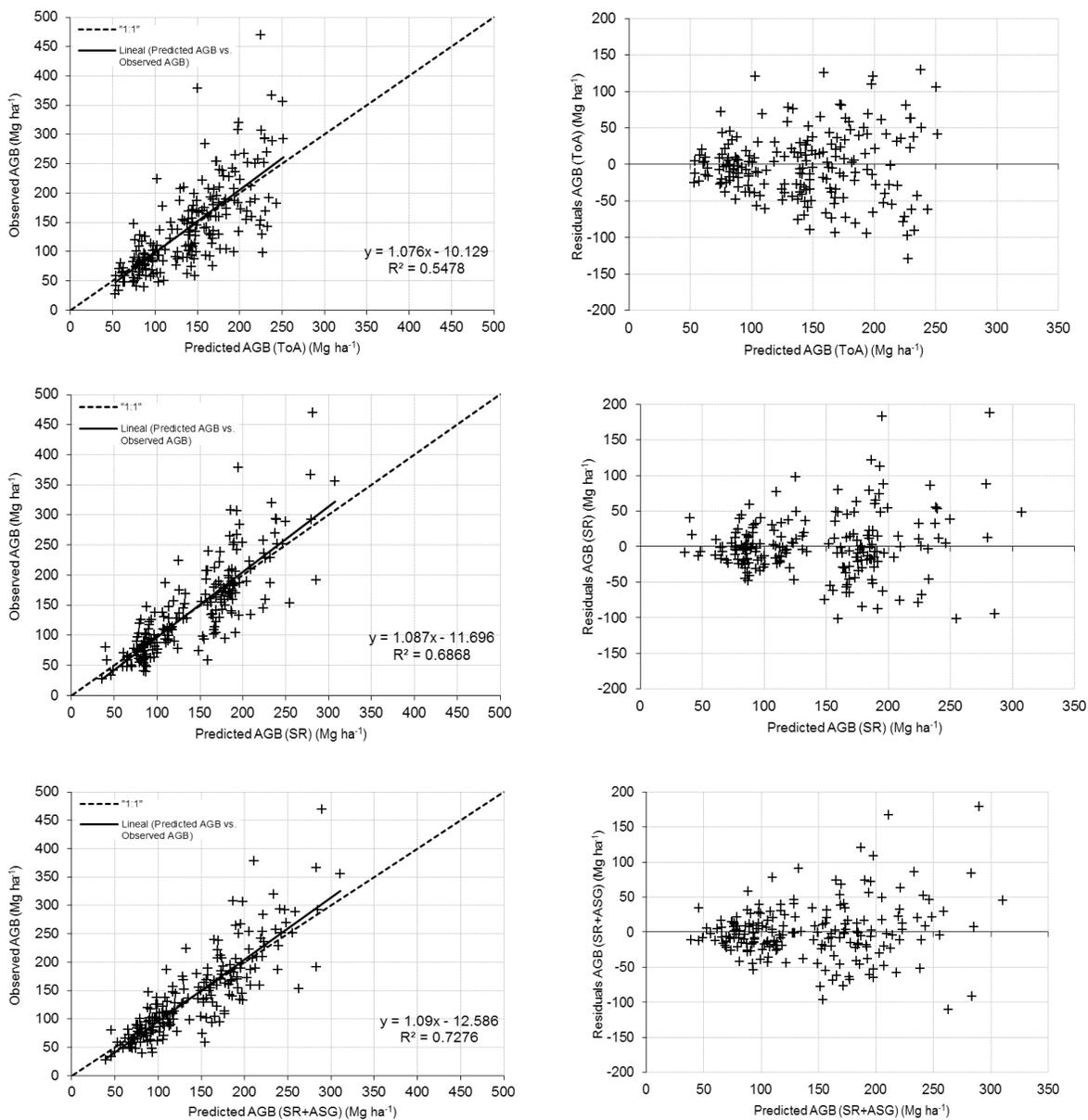
Figure 3. Graphs showing the distribution of the residuals and of the observed AGB values with ToA (upper), SR (middle) and SR with ASG variables (bottom).
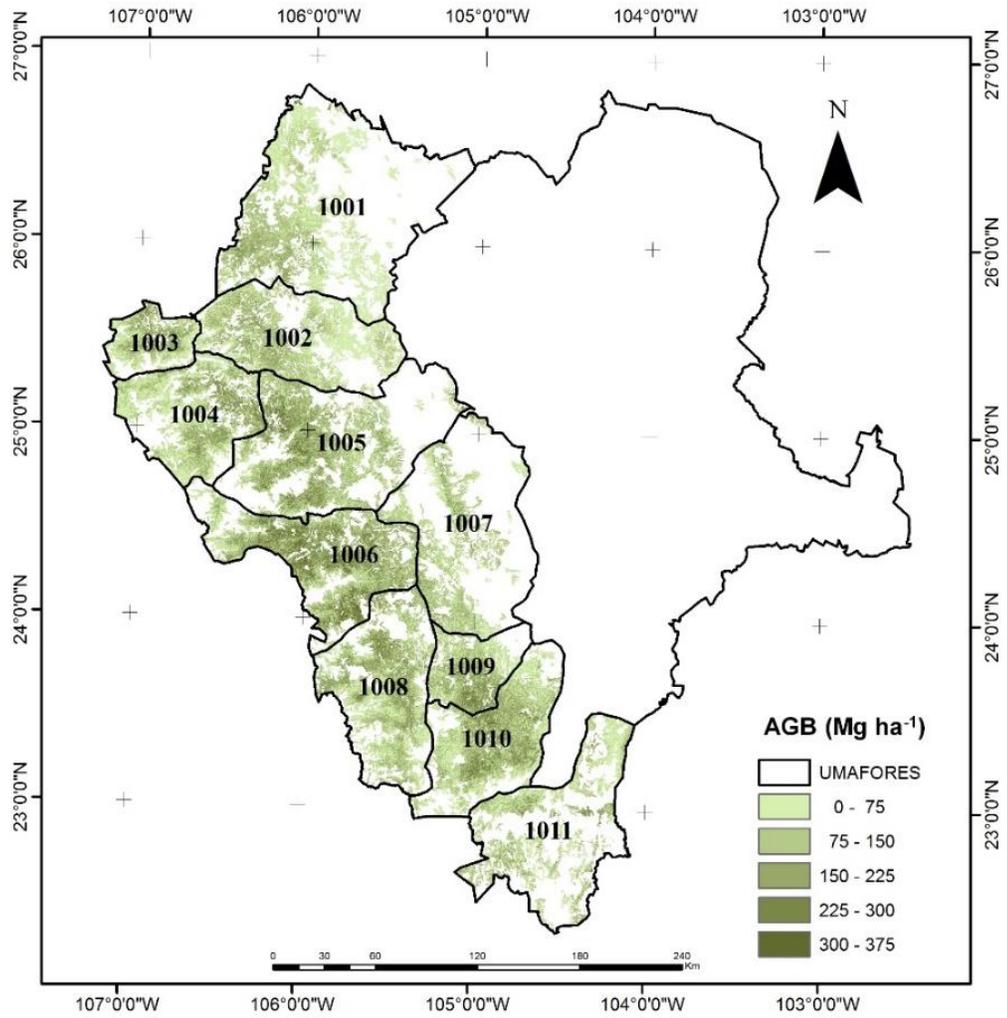
Figure 4. Spatial distribution of the total AGB in the SMO, state of Durango, Mexico.

Table 3. Estimation of AGB for the regional forest management units in the SMO, state of Durango, Mexico.

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

Discussion

The results of the present study demonstrate that the data acquired by a medium spatial resolution (Landsat) sensor are potentially useful for estimating AGB in structurally complex forests, such as those in the SMO in the state of Durango (Mexico), with satisfactory results and low cost. The deterministic predictors were the bands belonging to the blue, green and most red, near and mid infrared spectral regions. This finding was similar to that reported by (Jakubauskas, 1996), who demonstrated that reflectance in the red and near infrared regions yielded good predictions for AGB estimation in forest zones of the Yellowstone National Park, USA. In the present study, the model tended to underestimate AGB values above approximately 250 Mg·ha⁻¹. This might possibly be due to the saturation of NDVI, which is the most influential variable in predicting biomass for high values. In this sense, several studies have similarly found that the NDVI loses its sensitivity to dense vegetation because of the saturation in red and near infrared

wavelength in measuring and monitoring plant growth, vegetation cover and biomass production from satellite data (Huete *et al*., 1997; Lu *et al*., 2004; Mutanga and Skidmore, 2004). Models fitted ToA, SR, and SR with ASG, respectively, showed an increasing capacity to overcome these NDVI saturation problems.

Furthermore, (Fassnacht *et al*., 1997) concluded that the vegetation indices or individual bands, which include one or more bands in the infrared spectrum, provide satisfactory descriptions of zones occupied by conifer or broadleaf species. Moreover, in the case of SR, which was better than ToA, the contribution of Band 1 in two terminal nodes of the M5P model is associated with the structural variability of the canopy (Palestina *et al*., 2015). Günlü *et al*. [2014] found that the reflectance from Landsat TM satellite Band 1 was the best predictor of AGB (R^2 = 0.465, RMSE = 91,836 t·ha^{-1}), given the structural conditions of the canopy and understory, as the reflectance from this band increased as the AGB increased. In the present study, the result of the M5P analysis with SR spectral bands (R^2 = 0.69, RMSE = 42.17 Mg·ha^{-1}) was higher to that reported by Houghton *et al*. (2007), who analyzed data from the MODIS sensor (resolution, 500 m) and forest inventory data using the non-parametric Random Forest (R^2 = 0.61) method to map forest biomass in Russia.

In a recent study, Tian *et al*. (2014) used the non-parametric *k-nearest neighbours* (k-NN) technique to produce an optimized model (R^2 = 0.59, RMSE = 24.92 ton·ha^{-1}) from Landsat-TM images of a sample of 133 plots, with topographic correction based on sun-canopy-sensor (SCS + C). Likewise, Hall, [30] used Landsat 5 (TM) images rectified by SCS+C radiometric correction and compared the performance of the k-NN method and support vector machine (SVM) method for estimating AGB. They found that k-NN performed better (R^2 = 0.54, RMSE = 26.62 ton·ha^{-1}) than SVM (R^2 = 0.51, RMSE = 27.45 ton·ha^{-1}).

In general, most previous studies report significant relationships between AGB and the reflectance values yielded by each sensor. The reliability was within the range reported in diverse research studies that estimate AGB from medium resolution spectral data, from Landsat and SPOT, which often yield R^2 values between 0.50 and 0.70 with absolute errors of the estimates of between 30 and 60 Mg·ha^{-1} (Hall *et al.,* 2006; Foody *et al*., 2001; Tomppo *et al*., 2002; Zheng *et al*., 2004; Chen *et al*., 2009; Castillo *et al*., 2010; Tian *et al*., 2012; Guoa *et al*., 2014). The present study also shows that incorporation of spectral data and tree abundance estimated by species group in mixed and uneven-aged forests (SR with ASG), such as the

SMO, can increase the level of estimation of the AGB ($R^2 = 0.73$, RMSE = 39.40 Mg·ha^{-1}). In this sense, previous studies have reported significant variations in forest biomass estimation between different ecological zones, tree species, ages, density and management types (Henry et al., 2011; De-Meiguel *et al.,* 2014; Zou *et al.,* 2015).

In other studies (Richter *et al.,* 2009; Hantson and Chuvieco, 2011; Balthazar *et al.,* 2012) several authors have concluded that the spectral data derived after atmospheric and topographic correction may improve the accuracy of the biomass estimation, irrespective of the statistical method used. As the areas being monitored are mountainous zones, the quality of the data is negatively affected by the reflectance between sunny and shaded slopes. Interactive parameter fitting in the topographical correction methods may improve the quality of the spectral data and of the AGB estimates (Hantson and Chuvieco, 2011; Balthazar *et al.,* 2012).

Conclusions

In the present study, we estimated the AGB in the SMO in the state of Durango, Mexico, using the M5P technique and the analysis of medium-resolution satellite-based multi-spectral data, and field data collected from a network of 201 SPIFyS.

The findings show that the M5P method is potentially useful for estimating forest biomass. Data from the infrared channel of the Landsat-5 TM sensor proved best for discriminating or predicting AGB.

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

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

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

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

References

González E.M.S., González, E.M., Tena, F.J.A., Ruacho, G.L., López-Enríquez, I.L. 2012. Vegetación de la Sierra Madre Occidental, México: una síntesis. Acta Botánica Mex. 100: 351–403.

Secretaría de Medio Ambiente y Recursos Naturales, Anuario Estadístico de la Producción Forestal 2011. 2011. México, pp. 11-24.

Fuchs, H., Magdon, P., Kleinn, K., Flessa, H. 2009. Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. Remote Sens. Environ. 113, 518–531.

Barajas, F.H. Comparación Entre Análisis Discriminante no-Métrico y Regresión Logística Multinomial. 2007. Tesis de Maestría, Facultad de Ciencias, Universidad Nacional de Colombia: Medellín, Colombia.

Karjalainen, M., Kankare, V., Vastaranta, M., Holopainen, M., Hyyppa, J. 2012. Prediction of plot-level forest variables using TerraSAR-X stereo SAR data. Remote Sens. Environ. 117: 338–347.

Quinlan, J.R. 1992. Learning with Continuous Classes. In 5th Australian Joint Conference on Artificial Intelligence, Word Scientific: Singapore. pp. 343–348.

SRNyMA-CONAFOR. 2007. Plan Estrátegico Forestal 2030, Gobierno del estado de Durango: Durango, México.

Zhao, X., Corral-Rivas, J.J., Zhang, C.,Temesgen, H., Gadow, K. 2014. Forest observational studies-an essential infrastructure for sustainable use of natural resources. Forest Ecosyst. doi:10.1186/2197-5620-1-8.

Wehenkel, C., Corral-Rivas, J.J., Hernández-Díaz, J.C., Gadow, K. 2011. Estimating Balanced Structure Areas in multi-species forests on the Sierra Madre Occidental, Mexico. Ann. For. Sci. 68: 385–394.

Corral Rivas, J., Vargas, B., Wehenkel, C., Aguirre, O., Álvarez, J., Rojo, A. 2009. Guía para el Establecimiento de Sitios de Inventario Periódico Forestal y de Suelos del Estado de Durango, Facultad de Ciencias Forestales. Universidad Juárez del Estado de Durango: Durango, México.

Vargas-Larreta B., González-Herrera, L., López-Sánchez, C.A., Corral-Rivas, J.J., López-Martínez, J.O., Aguirre-Calderón, C.G., Álvarez-González, J.J. 2015. Biomass equations and Carbon content of the temperate forests of northwestern México. Biomass Bioenerg. in press.

United States Geological Survey. Available online: http://glovis.usgs.gov (accessed on 11 January 2011).

NASA. Landsat 7 Science Data Users Handbook, 2011. Available online: http://landsathandbook.gsfc.nasa.gov/pdfs/Landsat7_Handbook.pdf (accessed on Day, Month, Year).

Greenle, P.T. 1993. Remote Sensing science applications in arid environments. Remote Sens. Environ. 23: 143–154.

Keys, R.G. 1981. Cubic convolution interpolation for digital image processing. IEEE Trans. Acoust. Speech Signal Process. 29: 1153–1160.

Eastman, J.R. 2012. IDRISI Selva., Guía para SIG y Procesamiento de Imágenes, Clark Labs Clark University: Worcester, MA, USA.

Eastman, J.R. 2012. IDRISI Selva, Clark University: Worcester, MA, USA.

Gasparri, N.I., Parmuchi, M.G., Bono, J., Karszenbaum, H., Montenegro, C.L. 2010. Assessing multi-temporal Landsat 7 ETM+ images for estimating above-ground biomass in subtropical dry forests of Argentina. J. Arid Environ. 74: 1262–1270.

Breiman, L., Friedman, J., Olshen R., Sone, C. 1984. Classification and Regression Trees, Wadsworth International Group: Belmont, CA.

Wang, Y., Witten, I.H. 1997. Induction of model trees for predicting continuous classes. In Proceedings of the Poster Papers of the 9th European Conference on Machine Learning, Hamilton, New Zealand, 23 October 1997.

Mutanga, O., Skidmore, A.K. 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. Int. J. Remote Sens. 25: 3999–4014.

Fassnacht, K.S., Gower, S.T., MacKenzie, M.D., Nordheim, E.V., Lillesand, T.M. 1997. Estimating the leaf area index of North Central Wisconsin forests using the Landsat Thematic Mapper. Remote Sens. Environ. 61: 229–245.

Günlü, A., Ercanli İ., Başkent E.Z., Çakır G. 2014. Estimating aboveground biomass using Landsat TM imagery: A case study of Anatolian Crimean pine forests in Turkey. Ann. For. Res. 57: 289–298.

Houghton, R.A., Butman, D., Bunn, A.G., Krankina, O.N., Schlesinger, P., Stone, T.A. 2007. Mapping Russian forest biomass with data from satellites and forest inventories. Environ. Res. Lett. 4: 1–7.

Tian, X., Zengyuan, L., Zhongbo, S., Erxue, C., Christiaan, T., Xin, L., Guo, Y., Longhui, L., Feilong, L. 2014. Estimating montane forest above-ground biomass in the upper reaches of the Heihe River Basin using Landsat-TM data. Int. J. Remote Sens. 35: 7339–7362.

Guoa, Y., Tianb, X., Lib, Z., Linga, F., Chenb, E., Yanb M., Lib, C. 2014. Comparison of estimating forest above-ground biomass over montane area by two non-parametric methods. ISPRS J. Photogramm. Remote Sens. 92: 137–146.

Foody, G.M., Cutler, M.E., Mcmorrow, J., Pelz, D., Tangki, H., Boyd, D.S., Douglas, I. 2001. Mapping the biomass of Bornean tropical rain forest from remotely sensed data Global. Ecol. Biogeogr. 10: 379–387.

Tomppo, E., Nilsson, M., Rosengren, M., Aalto, P., Kennedy, P. 2002. Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass. Remote Sens. Environ. 82: 156–171.

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., le Moine, J., Ryu, S.R. 2004. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. Remote Sens. Environ. 93: 402–411.

Chen, W., Blain, D., Li, J., Keohler, K., Fraser, R., Zhang, K., Leblanc, S., Olthof, I., Wang, J., McGovern, M. 2009. Biomass measurements and relationships with Landsat-7/ETM+ and JERS-1/SAR data over Canada's western sub-arctic and low arctic. Int. J. Remote Sens. 30: 2355–2376.

Tian X., Zhongbo, S., Erxue, C., Zengyuan, L., Christiaan, T., Jianping, G., Qisheng, H. 2012. Estimation of forest above-ground biomass using multi-parameter remote sensing data over a cold and arid area. Int. J. Appl. Earth Obs. Geoinform. 14: 160–168.

Henry, M., Picard, N., Trotta, C., Manlay, R.J., Valentini, R., Bernoux, M., Saint-André, L. 2011. Estimating tree biomass of sub-Saharan African forests: A review of available allometric equations. Silva. Fenn. 45: 477–569.

De-Miguel, S., Pukkala, T., Assaf, N., Shater, Z. 2014. Intra-specific differences in allometric equations for aboveground biomass of eastern Mediterranean Pinus. Brutia. Ann. For. Sci. 71: 101–112.

Richter, R., Kellenberger, T., Kaufmann, H. 2009. Comparison of topographic correction methods. Remote. Sens. 3: 184–196.

Jakubauskas, M.E. 1996. Thematic Mapper characterization of lodge pole pine serals in Yellowstone National Park, USA. Remote Sens. Environ. 56: 118–132, doi:10.1016/0034-4257(95)00228-6.

Huete, A.R., Liu, H., van Leeuwen, W.J. 1997. The Use of Vegetation Indices in Forested Regions: Issues of Linearity and Saturation. In Proceedings of 1997 IEEE International Geoscience and Remote Sensing Symposium, Singapore, 3–8 August 1997, European Space Agency Publications Division: Noordwijk, The Netherlands. pp. 1966–1968.

Hall, M. 1999. Correlation-based Feature Selection for Machine Learning. Ph.D. Dissertation, University of Waikato, Hamilton, NewZealand.

Foody, G.M., Boyd, D.S., Cutler, M.E.J. 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. Remote Sens. Environ. 85: 463–474.

Gibbons, J.D., Chakraborti, S. 2003. Nonparametric Statistical Interference, Marcel Denker, Inc.: New York, NY, USA.pp. 645.

Sánchez, O., Vega, E., Peters, E.Y., Monrroy V.O. 2003. Conservación de Ecosistemas Templados de Montaña en México, Instituto Nacional de Ecología (INE–SEMARNAT): México D.F., México.

Lu, D., Mausel, P., Brondızio, E., Moran, E. 2004. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. For. Ecol. Manag. 198: 149–167.

Muukkonen, P., J. 2005. Heiskanen. Estimating biomass for boreal forests using ASTER satellite data Combined with standwise forest inventory data. Remote Sens. Environ. 99: 434–447.

Hall, R.J., Skakun, R.S., Arsenault, E.J., Case, B.S. 2006. Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. For. Ecol. Manag. 225: 378–390.

Chuvieco, E. 2010. Teledetección Ambiental: La observación de la tierra desde el espacio, Ariel.: Barcelona, España.

Castillo Santiago, M.A., Ricker, M., Jong, B.H.J. 2010. Estimation of tropical forest structure from SPOT-5 satellite images. Int. J. Remote Sens. 31: 2767–2782.

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D. 2010. Detecting trend and seasonal changes in satellite image time series. Remote Sens. Environ. 114: 106–115.

Hantson, S., Chuvieco, E. 2011. Evaluation of different topographic correction methods for Landsat imagery. Int. J. Appl. Earth Obs. Geoinform. 5: 691–700.

Balthazar, V., Vanacker, V., Lambin, E.F. 2012. Evaluation and parameterization of ATCOR3 topographic correction method for forest cover mapping in mountain areas. Int. J. Appl. Earth Obs. Geoinform. 18: 436–450.

Liang, S., Li, X., Wang, J. 2012. Advanced Remote Sensing: Terrestrial Information Extraction and Applications, Academic Press: Oxford, UK.

ArcGIS for Desktop, version 10. 2012. software for analyzing geographic information systems, ESRI: Redlands, CA, USA.

Palestina, R.A., Equihua, M., Pérez-Maqueo, O.M. 2015. Influencia de la complejidad estructural del dosel en la reflectancia de datos Landsat TM. Madera y Bosques. 21: 63–75.

Zou, W.-T., Zeng, W.-S., Zhang, L.-J., Zeng, M. 2015. Modeling Crown Biomass for Four Pine Species in China. Forests. 6: 433–449.

Zhu, X., Liu, D. 2015. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. ISPRS J. Photogramm. Remote Sens. 102: 222–231.

CAPÍTULO 4. A COMPARISON OF MACHINE LEARNING TECHNIQUES APPLIED TO LANDSAT-5 TM SPECTRAL DATA FOR BIOMASS ESTIMATION

Pablito M. López-Serrano [1], Carlos A. López-Sánchez [1], Juan G. Álvarez-González [2] and Jorge García-Gutiérrez [3]

Abstract

Machine learning combines inductive and automated techniques for recognising patterns. These techniques can be used with remote sensing datasets to map AGB with an acceptable degree of accuracy for evaluation and management of forest ecosystems. Unfortunately, statistically rigorous comparisons of machine learning algorithms are scarce. The aim of this study was to compare the performance of three most common non-parametric machine learning techniques reported in the literature (Support Vector Machine [SVM], k-nearest neighbor [kNN] and Random Forest [RF]) with that of the parametric multiple linear regression (MLR) technique, for estimating AGB from Landsat-5 Thematic Mapper (TM) spectral reflectance data, texture features derived from the normalized difference vegetation index (NDVI) and topographical features derived from a digital elevation model (DEM). The results obtained for 99 permanent sites (for calibration/validation of the models) established during the winter of 2011 by systematic sampling in the state of Durango (Mexico) showed that SVM performed best once the parameterization had been optimized. Otherwise, SVM could be outperformed by RF. However, the kNN yielded the best overall results in relation to the goodness-of-fit measures. The findings confirm that non-parametric machine learning algorithms are powerful and precise regression tools for estimating AGB with datasets derived from sensors with medium spatial resolution.

[1]Instituto de Silvicultura e Industria de la Madera, Universidad Juárez del Estado de Durango,

[2]Escuela Politécnica Superior, University of Santiago de Compostela, Campus universitario s/n, Lugo, Spain

[3]Departamento de Lenguajes y Sistemas Informáticos. Universidad de Sevilla [*]Autor para correspondencia: jorgarcia@us.us

Introduction

Forest biomass plays an important role in the global climate system as forest ecosystems absorb approximately one twelfth of the Earth's atmospheric carbon stocks every year (Malhi *et al*., 2002), and much of this carbon is stored as aboveground biomass (AGB). The importance of forest biomass has been underlined by the United Nations Framework Convention on Climate Change (UNFCCC), which has identified AGB as an Essential Climate Variable (GCOS, 2010). Moreover, quantification of AGB and modelling of the associated dynamics are important to support decision-making models in different fields, including energy and materials provision for human use (FAO, 2001, 2006), forest fragmentation (e.g. Malhi and Phillips, 2004) and biodiversity conservation (e.g. Bunker *et al*., 2005). Accurate monitoring of forest biomass and how it changes at local to global scales is therefore of critical importance towards a better understanding of these processes (Lu, 2006; Hartig *et al.,* 2012; Le Toan and Quegan, 2015).

The most accurate method of estimating forest biomass is based on field measurements, however, estimating biomass in large areas is not an easy task and is hindered by the high costs (both time and money) associated with fieldwork (Lu *et al.,* 2016).

Remote sensing has been shown to be a practical option that helps to overcome these limitations as it enables to get information of forest in large areas with reasonable effort. This is now the primary data source for large-scale biomass estimation (e.g. Andersen *et al.,* 2011, Lu et al., 2016). Over the past few decades, the so-called passive sensors (i.e. sensors that use the solar radiation reflected or emitted by the objects detected at the earth's surface) have been used to estimate AGB (e.g. Lu *et al.,* 2012; Frazier *et al.,* 2014). Considering the advantages and limitations of different remote sensing images, the medium-resolution (pixel size, 30 m) Landsat-5 TM sensor is one of the most widely used for biomass estimation (e.g. Agarwal *et al.,* 2014; Pflugmacher *et al.,* 2014, Dube and Mutanga, 2015, Zhu and Liu, 2015). The advantages of using the Landsat-5 TM sensor are that numerous historical spatio-temporal archives are available (images since 1972) and the Landsat data is free of economic cost for users than high resolution sensors, particularly for analysis of large areas. For a review of Landsat imagery-based AGB estimations, see Wu *et al.* (2016).

Independently of the type of sensor used, model accuracy and error estimation vary in relation to a series of factors such as the structure of the field data and the statistical

techniques used (Ghosh *et al.,* 2014). The most common model used in estimating forest biomass from remote sensing data is the regression-based model (e.g. Tian *et al.* 2012; Lu

2012; Næsset et al. 2013), however, the accuracy of estimates obtained with small numbers of sample plots or when there is a weak linear relationship between variables and biomass is rather low (Lu *et al.,* 2016). Non-parametric modelling approaches, which make no assumptions about the statistical distributions of the original data and relationships between predictor and response variables, have also been used to relate AGB and remotely sensed features. Various recent studies have explored the use of non-parametric approaches for estimating AGB with remote sensing data (e.g. Breidenbach *et al.,* 2012; Mutanga *et al.,* 2012; Jung *et al.,* 2013; Fassnacht *et al.,* 2014).

Machine learning involves different techniques (mainly non-parametric) that focus on automated and inductive learning to recognise patterns (Cracknell and Reading, 2014) in data (e.g. patterns in remote sensing data related to AGB in a set of located plots), once the pattern is learned, it can be applied to yield a prediction or classification in areas where it is not possible to carry out fieldwork to quantify an objective variable (e.g. AGB). In the last decade, various machine learning techniques such as Support Vector Machine (SVM), k-nearest neighbour (kNN) and Random Forest (RF) have been used to develop predictive models of AGB in large areas. Thus, Shataee (2013) showed that kNN performed better than SVMs, RF and Artificial Neural Networks (ANN) for estimating biophysical variables such as basal area. More recently, Garcia-Gutierrez *et al.* (2015) showed that SVM models performed best for estimating forest variables from Light Detection and Ranging (LIDAR), while Wang *et al.* (2016) showed that RF outperformed SVM and ANN for estimating wheat biomass from remote sensing data. For a more complete review of research being carried out to retrieve vegetation biomass from remote sensing data using machine learning methods, see Ali *et al.* (2016).

The goodness-of-fit of models derived from spectral data is usually evaluated by the coefficient of determination (R^2) and the root mean square error (*RMSE*). These measures report the performance of the model in predicting the data used to fit the model, however, because the quality of the fit does not necessarily reflect the quality of the prediction, assessment of their validity is often needed to ensure that the predictions represent the most likely outcome in the real world (Yang *et al.,* 2004). The only method that can be regarded as

''true'' validation involves the use of a new independent data set (Pretzsch *et al.,* 2002; Yang *et al.,* 2004), however, the scarcity of such data forces the use of alternative approaches, such as Cross Validation (CV) to enable evaluation of the quality of a particular fitting technique and minimize the risk of overfitting (Molinaro *et al.,* 2005). Unfortunately, most studies involving estimation of AGB do not use CV as part of the model development.

For rigorous comparison of the performance of different machine learning techniques, the study should also be accompanied by statistical validation of the results within a statistical framework (i.e. not merely calculating statistics such as R^2 or *RMSE*). Although this is well known in the field of machine learning (García *et al.,* 2010), this type of validation is not common in remote sensing, even though machine learning plays an important role in many biomass estimation studies. This fact may have led to some degree of discordance in the scientific literature, in which we can find examples of kNN, SVM and RF outperforming each other (Shataee 2013; Garcia-Gutierrez *et al.,* 2015; Wang et al., 2016).

The objective of this study was to analyze and statistically compare the performance of three non-parametric techniques (SVM, kNN and RF) and the parametric Multiple Linear Regression (MLR) technique for estimating AGB. The techniques were tested with Landsat-5 TM surface spectral reflectance data, texture features derived from the normalized difference vegetation index (NDVI) and topographical features derived from a digital elevation model (DEM) in the Sierra Madre Occidental (state of Durango, Mexico). The results obtained with each technique were compared after application of CV and posterior statistical validation of the mean rankings obtained for each.

Materials and Methods

Study area

The study site is located in the Sierra Madre Occidental, in the north of the state of Durango (Mexico), and covers an area of 1,142,916 ha (Figure 1). The climate is humid temperate, with rainfall in summer (relative humidity, 50.1%). The average temperature ranges from 8 to 20 °C and the annual precipitation from 400 to 1200 mm. The average altitude above sea level in this area is 1,900 m. 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). The forests are basically mixed and uneven-aged pine-oak stands, with a canopy cover ranging from 32 to 100%. These forests

have been subject to selective harvesting for almost a century to provide a mixture of services to local communities. This structure is the result of the management history, which has depended on land ownership and the economic and social changes that have taken place in the state, as well as natural conditions (Wehenkel *et al.,* 2011).
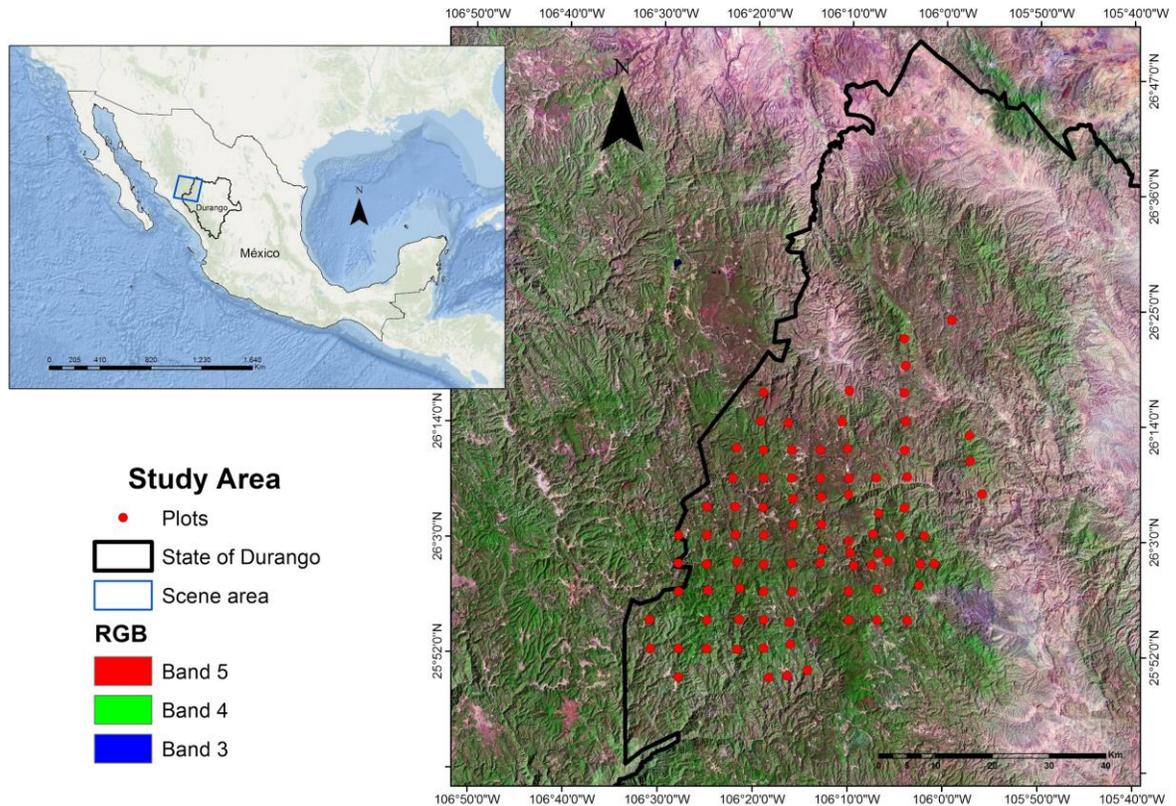
Figure 1. Geographical location of the study site and sample plots used in the study.

Dataset

Field data

A network of 99 permanent sampling plots (Sitios Permanentes de InvestigaciÃs¸n Forestal y de Suelos, SPIFyS) 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) of a grid of equidistant points separated by three

or five km, depending on the orography of the study area. In each plot (squares of side 50 m), all species of trees were recorded and the diameter at breast height (cm) and total height (m) of all standing trees were measured. Species-specific individual tree models developed by Vargas-Larreta (2013), were used to estimate the total AGB of field plots by tree value aggregation. The R2 and the RMSE of the models used ranged from 0.87-0.99 and 22.8-95.2 kg respectively.

Summary statistics including number of observations, mean, standard deviation, minimum and maximum values of AGB per hectare are summarised in Table 1.

Table 1. Total biomass statistics (expressed in Mg ha^{-1})

No. of observations	Mean	Standard deviation	Minimum value	Maximum value
99	89.03	43.45	2.70	234.03

Spectral data

The spectral data were derived from a satellite image Landsat-5 TM obtained in April 2011 (path 32, row 42) and covering the entire study area (available from the US Geological Service webpage, at http://glovis.usgs.gov/). Landsat-5 TM data have a spatial resolution of 30m with a revisit period of 16 days. Bands 1, 2, 3, 4, 5 and 7 (level L1T) of Landsat-5 TM were used in the present study, band 6 was not used, because of its thermal characteristics, its coarse spatial resolution (120 m) and the low contrast in the forest area (NASA, 2011). The satellite images were radiometrically, atmospherically and topographically corrected using the ATCOR3$^{®}$ module, regarded as particularly suitable for mountainous zones. The ATCOR3$^{®}$ module (Geosystems 2013) first calculates the radiance at sensor level (W sr^{-1} m^{-2}) from the image pixel. 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 the characteristics 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 vapour

bands were available, dehazing/cloud removal and atmospheric water retrieval settings were kept as "default", which, in this case, is recommended by the ATCOR3® User Manual (Geosystems, 2013). The corrections were implemented with the ERDAS® IMAGINE® 2013 software (ERDAS, Inc. 2014). A number of vegetation indices were computed from the atmospherically and topographically corrected image bands and included in the biomass estimation models for evaluation as possible regressor features (Table 2).

Texture parameters

The texture features homogeneity, contrast, dissimilarity, mean, standard deviation, entropy, second order angular moment and correlation (Haralick *et al.,* 1973) were calculated from the NDVI image based on grey level co-occurrence matrices, with the aim of including information combining the spatial and spectral domain of the remotely sensed imagery in the biomass estimation models. We used NDVI texture features rather than each spectral band of Landsat-5 TM to avoid saturating high biomass values (Mutanga and Skidmore, 2004). As it also becomes more difficult to obtain an optimal subset as the number of attributes increases, we therefore aimed for a compromise between quantity and quality. The features were calculated using PCI Geomatica2013® software (PCI Geomatics Inc., 2013) and three different scales of operation were considered by using moving window sizes of 3x3, 5x5 and 7x7 pixels (Table 2).

Terrain variables

Terrain features are directly related to forest species composition, tree height growth and other forest stand variables, enabling these to be modelled (McNab, 1989, Roberts and Cooper, 1989). First and second order terrain features were therefore derived from the 5x5 low pass filtered Digital Elevation Model (DEM) of the study area with a spatial resolution of 15 m. The DEM was derived from LIDAR data and corresponds to an array of elevation data interpolated to 15 m resolution from the coordinates of the last return of the pulses emitted (INEGI, 2014). The final set of features derived from Landsat-5 TM sensor and from the DEM, which were used as possible predictors (independent variables) for estimating AGB (which played the role of dependent variable), are shown in Table 2.

Finally, the sample plots were geopositioned with the aim of extracting the pixel value average with an associated buffer of 25 m for each described feature, to obtain a database with the

mean biomass values and the associated features for each plot. The extraction was carried out using R statistical software (R Core Team, 2014) and the "raster" package.

Comparison framework

Machine learning techniques

Three non-parametric machine learning techniques and one parametric technique were applied to data from the study area in order to compare their performance: (i) k-Nearest Neighbour (kNN), (ii) Support Vector Machine (SVM), (iii) Random Forest (RF) and (iv) Multiple Linear Regression (MLR). All these techniques were used to estimate AGB using as possible predictors the variables included in Table 2.

The parametric MLR technique is the most commonly used in this kind of study (Fassnacht *et al.,* 2014). Moreover, this type of model is easy to understand and is widely used in most scientific disciplines. However, unlike the non-parametric approaches, MLR relies on certain assumptions, such as the fundamental least squares assumption of independence and equal distribution of errors with zero mean and constant variance, which can be violated by factors such as non-normality of variables, multicollinearity of variables and heteroscedasticity of error variance.

Nearest Neighbour (NN), a well-known machine learning technique used in remote sensing (Shataee, 2013), makes a prediction by using the information about the neighbours of the instance to be regressed (Cover and Hart, 1967). The NN depends on a parameter, usually called k, which determines the number of neighbours used by the algorithm. The technique is therefore usually called kNN when more than one neighbour is used. Although the idea behind this type of technique is quite intuitive, the resulting model is not easy to interpret as all results depend on a training set itself.

Table 2. Variables used in machine learning for biomass estimation

	Variable	Reference
Vegetation Index		
NDVI	Normalized Difference Vegetation Index	Rouse et al. (1974)
MSAVI2	Modified Soil-adjusted Vegetation Index	Qi et al. (1994)
SAVI	Adjusted Soil Vegetation Index	Huete (1988)
IAF	Leaf Area Index	Baret and Guyot (1991)
ALB	Albedo	Asrar (1989)
Fpar	Fraction of Photosynthetically Active Radiation	Asrar et al. (1984)
FSR	Flow Solar Radiation	Brutsaerts (1975)
Texture (NDVI)		
HOL	Homogeneity	
CO	Contrast	
DI	Dissimilarity	
ME	Mean	Haralick et al. (1973)
SDT	Standard Deviation	
EN	Entropy	
ASM	Angular Second Moment	
CR	Correlation	
Terrain (DEM)		
Altitude	Altitude	
B	Slope	
TRASP	Transformed Aspect	Roberts and Cooper (1989)
TSI	Terrain Shape Index	McNab (1989)
WI	Wetness Index	Moore and Nieber (1989)
PC	Profile curvature	
PLC	Plan curvature	Wilson and Gallant (2000)
C	Curvature	

SVMs have been developed from artificial neural networks (Cortes and Vapnik, 1995) and have been used in many scientific fields (e.g. Abedi *et al.,* 2012, Bayoudh *et al.,* 2015, Garcia-Gutierrez *et al.,* 2015). SVM models are developed by a set of vectors (or hyperplanes if greater dimension is requested) that separate instances of different labels (classification) or minimize the mean error (regression). Kernel functions are used to overcome the limitations

associated with linear separability in SVM models. Appropriate selection of the kernel function and the kernel regularization parameters is important in relation to the SVM model behaviour, which can make this type of technique more difficult to implement for users. As with kNN, the models produced using SVM are more difficult to interpret than those of MLR. RF is not exactly a classification or regression technique, but a combination of other techniques, mainly regression or classification trees (Breiman, 2001). The success of this technique is based on the use of numerous trees developed with different independent variables that are randomly selected from the complete original set of features (e.g. Deschamps *et al.,* 2012, Wang *et al.,* 2016). The number of predictors used by trees and the number of trees are established by the users.

WEKA open source software (Hall *et al.,* 2009) was used to implement all of the techniques compared. Thus, *Linear Regression* was used for MLR, *IBk* for kNN, *SMOreg* with Polynomial and Gaussian kernels for SVM and an adaptation of the RF implementation of WEKA for regression (using M5P as the basic regression technique for the development of this ensemble).

Feature selection, parameterization and validation

In machine learning, spurious data features must be removed before a model is generated (Hall, 1999). Thus, the variables that are potentially most important are selected. Some techniques (e.g. SVM and RF) carry out this selection, but others may be seriously affected by excessively large combinations of variables (e.g. the Hughes effect [Hughes, 1968] in kNN and multicollinearity in MLR). This is a common situation in this type of analysis because of the large set of predictor variables that can be calculated from remote sensing data (Packalén *et al.,* 2012). Moreover, correct functioning of different machine learning techniques depends on a proper parameterization (set-up of their parameters, i.e. variables that modify the behaviour of the machine learning techniques). In this study, both of these steps (feature selection and parameterization) were carried out via a metaheuristic search (Samadzadegan *et al.,* 2012). From the possible metaheuristic techniques (i.e. a method of optimization that provides a near-optimal solution in computationally affordable time), we selected an evolutionary algorithm which is illustrated in Figure 2. The algorithm starts with a population of random solutions (Initial Population in Figure 2) called individuals and ranks them

according to fitness of the individuals (Fitness Sorting in Figure 2). In the present study, the fitness was evaluated by the *RMSE* obtained with a training set. A new population of individuals is then created by mating parents (random selection of coefficients shown in Figure 2), selected with a probability proportional to their fitness, and later mutating the new individuals with a given probability (in this case, a value will be randomly selected and changed to a new random value as can be seen in Figure 2) .

The general scheme described in Figure 2 was modified slightly according to the specific regression technique. Thus, we used a specific design for MLR (see García-Gutiérrez et al., 2014) and an adaptation of the genetic algorithm of Huang and Wang (2006) for the non-parametric techniques (kNN, SVM and RF). In the kNN method, pure selection (coefficients associated with each feature as 1 or 0 depending on whether the predictor is selected or not) was substituted by weighting each attribute (real value between 0.0 and 1.0), which enables better adaptation of the algorithm to the characteristics of kNN (see Mateos *et al.,* 2012). In SVMs, the type of kernel is another parameter to be optimized and had two possible values (Radial Basis Function and Polynomial). The parameters optimized for each machine learning technique are included in Table 3.

Table 3. Intervals used by the evolutionary algorithm to search for the different optimal parameters.

	Technique	Name	Minimum	Maximum
kNN		k	1	20
SVM		EPSILON	0	0.2
SVM	Gaussian-kernel-only	GAMMA	0.01	2.0
SVM	Polynomial-kernel-only	EXP	1	5
SVM	Polynomial-kernel-only	C	1	100
RF		NT	1	100
RF		NF	1	5

Where:

K= Number of neighbours

C= Penalty factor imposed in SVM per instance of misclassification in training.

NT=Number of trees that form each ensemble.

NF= Number of attributes selected for constructing each tree
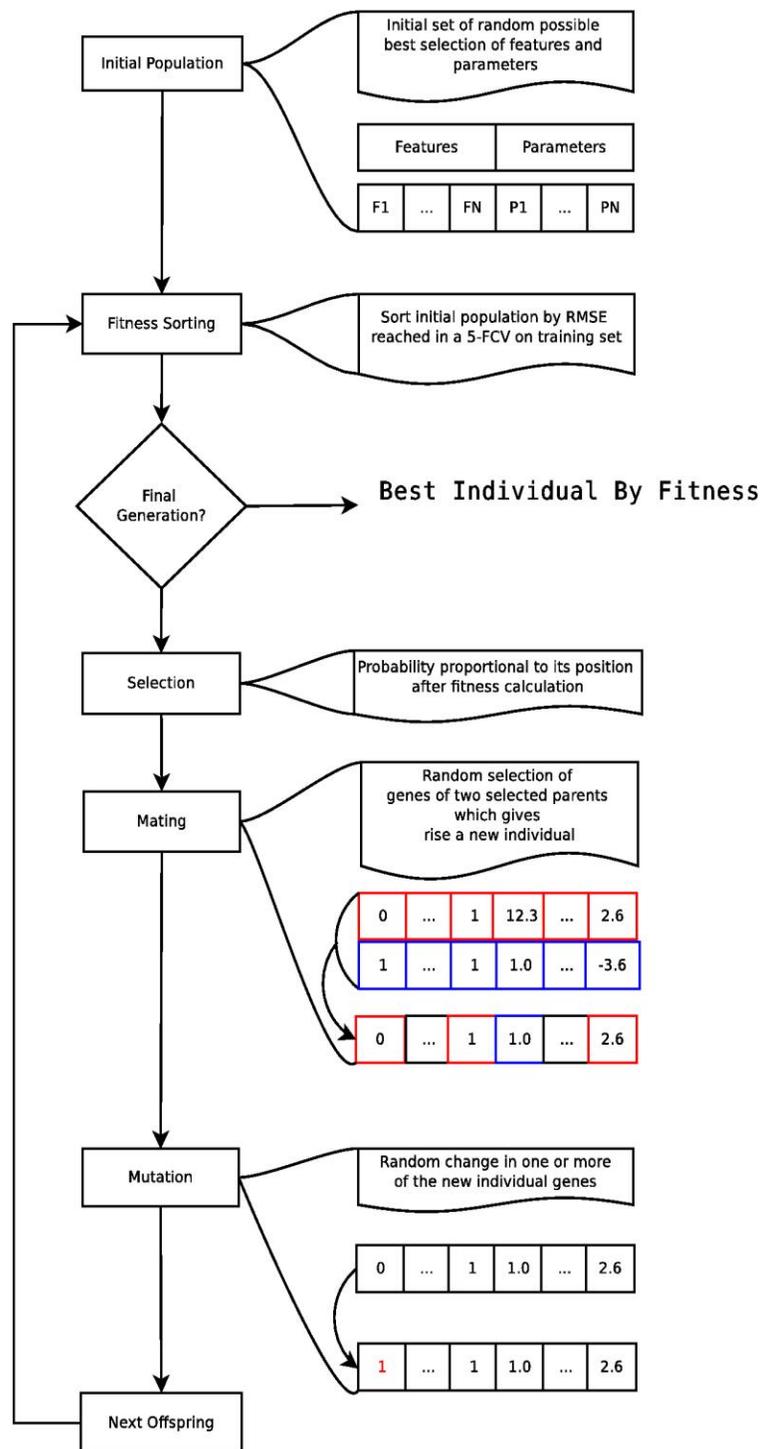
Figure 2. Description of the evolutionary procedure used to determine the best methods for parameterization and feature selection.

For comparison of the different techniques, validation was based on the Leave-One-Out CV technique. This is a special case of k-fold CV in which k is equal to the number of observations and a prediction is obtained as many times as there are observations in the dataset (Packalén *et al.,* 2012). In other words, an observation is excluded (target observation), and a prediction is computed with the other observations (reference observations). The prediction can be evaluated by the target observation. This procedure is repeated for every single observation. The final quality of a technique evaluated with CV is based on the averaged error obtained. A general description of the procedure is provided in Figure 3.
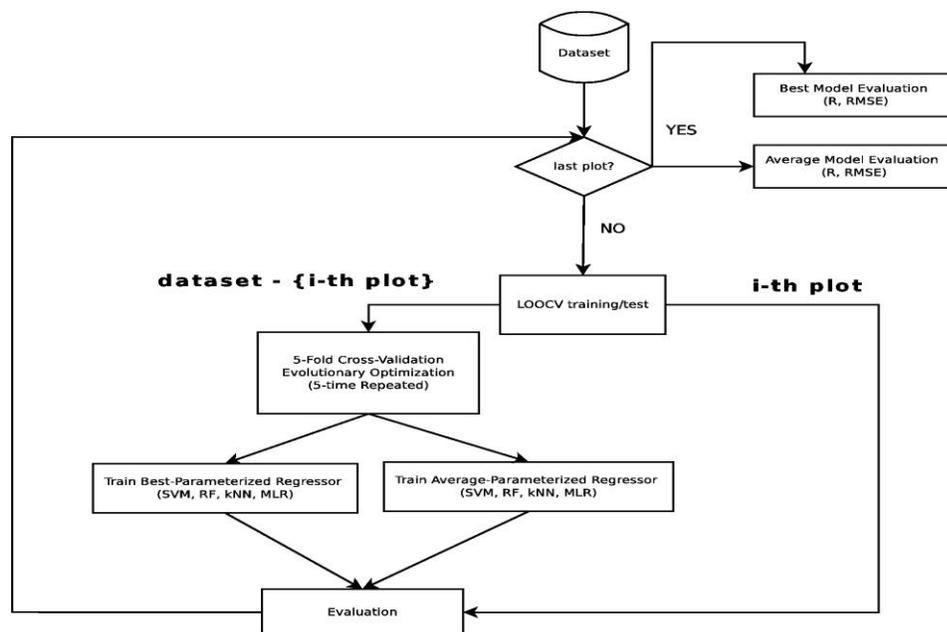
Figure 3. Description of the LOOCV evaluation of the techniques compared in the text.

Parameterization of each submodel at the different stages of the CV was repeated 5 times for each technique to prevent skew (due to the random nature of the evolutionary algorithms applied to predictor selection and parameterization). The best submodel and the average submodel for the 5 executions, ranked in terms of the *RMSE* reached in the evolutionary procedure, were used to calculate the goodness-of-fit statistics.

Statistical analysis

The error of the predictions in the CV was compared for each technique in terms of R^2 and *RMSE*. In addition, for statistical analysis of differences between the methods, the absolute errors of the predictions made by each technique throughout the 99 iterations in the CV were compared (the number of iterations is equal to the number of instances in the database, which in this case refers to the 99 plots available). In theory, this should be carried out by Analysis of Variance (ANOVA), if the data comply with the underlying assumptions of independence, normality and homoscedasticity required for parametric tests. These conditions can be tested by respectively the Shapiro-Wilk test, Lilliefor's test and Levenes' test. If the data do not comply with these conditions, a non-parametric test such as the Friedman's (aligned) test (described by García *et al.,* 2010) should be used.

Friedman's (aligned) test first obtains the mean ranking for each technique by taking into account the position obtained for each of the results relative to the other. Thus, a ranking of 1 for one of the techniques signifies that the result is the best of all results obtained in the procedure, whereas a rank of $m * n$, where m is the number of techniques being compared and n is the number of tests, indicates the poorest result obtained. After establishing the mean rankings for each technique, the Friedman's (aligned) test and Holm's post hoc procedure (see Luengo *et al.,* 2009 and García *et al.,* 2010, for a complete description) are used for statistical validation of the differences between the methods compared.

Results and discussion

The results indicated that the features for estimating AGB by the different machine learning algorithms evaluated can be classified into three different groups. In order of decreasing importance, the first group comprises the spectral bands and the spectral indices, the second group comprises the first and second order terrain topographical variables derived from DEM, and the third group comprises the texture features derived from the NDVI. The correlation derived from the texture image with a moving window of 7x7 pixels (CR7x7, see Table 2 for acronyms and abbreviations) was also a key feature in the MLR technique, although it was not important in the other techniques (see Figures 4 and 5).
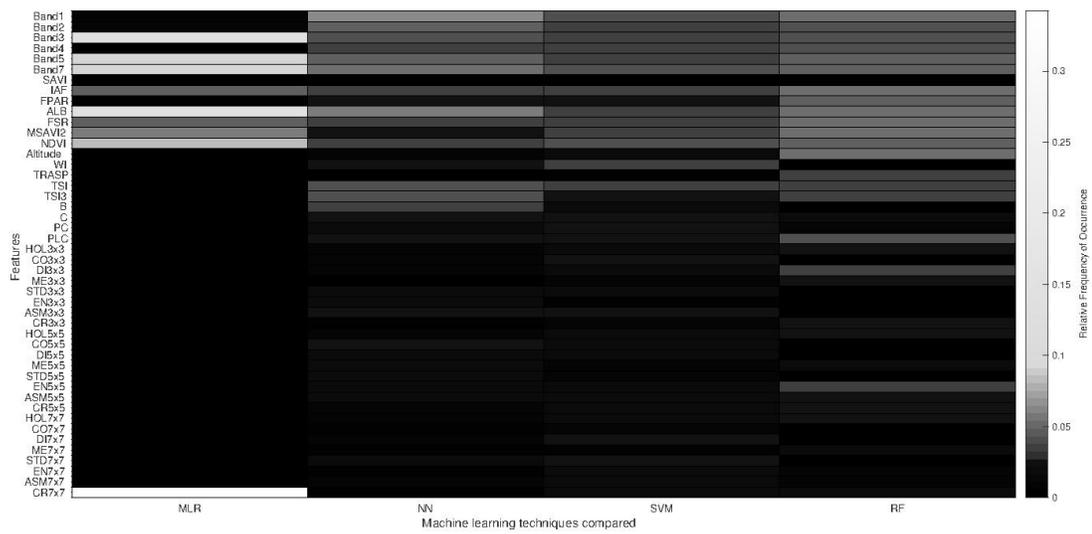
Figure 4. Relative frequency of ocurrence (importance) of each attribute in the best models obtained by each technique (in terms of the sum of residuals).
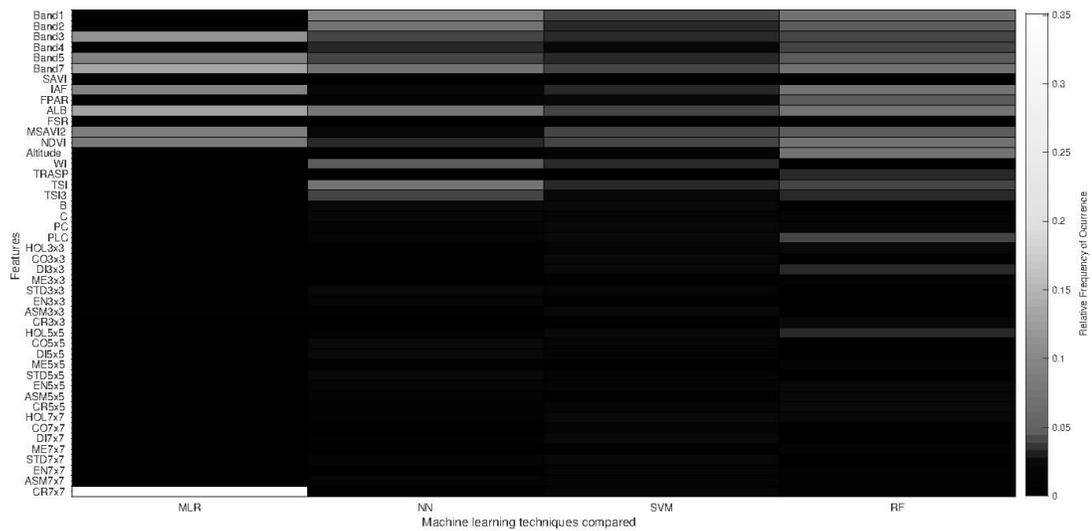
Figure 5. Relative frequency of ocurrence (importance) in the averaged models obtained by each technique (in terms of the sum of residuals).

The results obtained in terms of the RMSE were used for statistical comparison of the techniques. The comparison is summarised in histograms showing the relative positions reached (rankings) for each technique (Figures 6 and 7). Qualitatively, the SVM technique yielded the best results when the parameterization and selection of predictors were relatively optimal, whereas on average, the RF technique produced the best results.
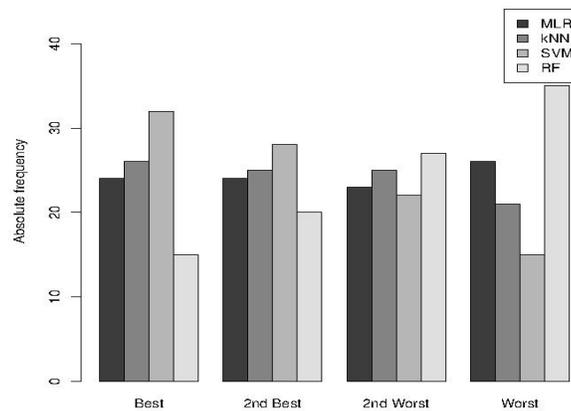
Figure 6. Absolute frequency of relative position achieved by each technique (ranking) with the best parameterization of 5 executions.
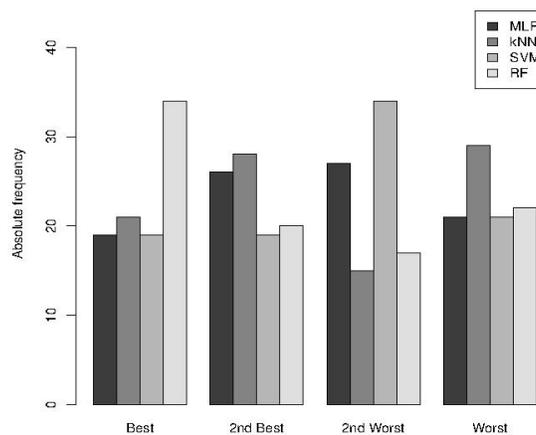
Figure 7. Absolute frequency of the relative position achieved (ranking) by each technique with average parameterization.

The rankings associated with the Friedman's (aligned) test and a post hoc Holm's test for paired comparison of the best algorithms (Table 4) confirm the idea (previously outlined in Figures 6 and 7) that SVM and RF techniques yielded the best results, considering respectively the best model and the averaged models for the 5 executions per plot. Friedman's (aligned) test yielded a p-value <0.0001, thus confirming rejection of the null hypothesis (i.e. that the overall performance of the methods was not significantly different).

Table 4. Mean rankings for the models obtained by each technique (the best mean ranks are indicated in bold).

Technique	Ranking	
	Best Models only	**Averaged Models**
MLR	197.98	188.11
Knn	186.69	184.37
RF	232.16	**176.04**
SVM	**161.18**	197.48

Application of Holm's procedure revealed that the results yielded by the SVM technique were significantly different from those produced by all other techniques except kNN (p=0.1131, higher than the significance levels of the test, α=0.05). Comparison of RF and the other techniques for the averaged models showed that none of the comparisons was statistically significant and it was therefore not possible to infer that RF performed better than the other models. The results of both procedures are summarised in Table 5.

Table 5. Results of the post hoc Holm's test of paired comparisons for SVM (best models only) and RF (averaged models). The comparisons that were not significantly different are indicated in bold type.

Best Models only				Averaged Models			
Technique	**P**	**z**	**Holm**	**Technique**	**p**	**z**	**Holm**
RF	0.000	4.408	0.0167	SVM	**0.174**	1.36	0.0167
MLR	0.022	2.285	0.025	MLR	**0.444**	0.77	0.025
kNN	**0.113**	1.584	0.05	kNN	**0.598**	0.53	0.05

Finally, the results for all plots were used to calculate the goodness-of-fit statistics: R^2 and *RMSE*. Table 6 summarises the application of these to the best models and the averaged models, in which kNN was the best technique in both cases. Maps of the AGB estimations obtained for the study area by each technique are shown in Figure 8.

Table 6. Summary of the goodness-of-fit statistics taking into account the overall results for the 99 plots. The best models are indicated in bold.

		MLR	kNN	SVM	RF
Best models only	R^2	0.54	**0.66**	0.62	0.48
	RMSE (Mg ha^{-1})	29.61	**26.64**	27.28	31.61
Averaged models	R^2	0.36	**0.41**	0.30	0.29
	RMSE (Mg ha^{-1})	34.67	**33.53**	36.15	39.20
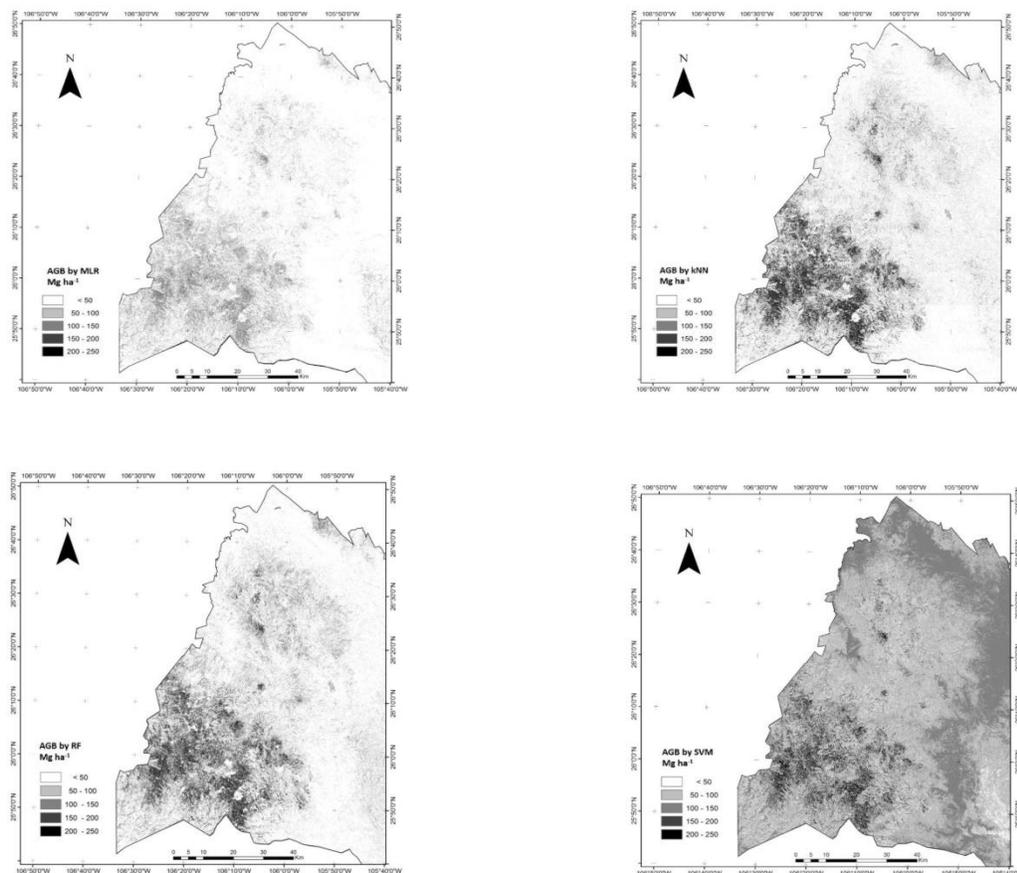
Figure 8. Biomass maps derived by each technique: MLR (top left), kNN (top right), RF (bottom left) and SVM (bottom right).

The results showed that the features that were most important for estimating AGB by the different machine learning techniques evaluated (kNN, RF and SVM) correspond to the bands

and spectral indices derived from the Landsat-5 TM sensor (Band 1, Band 5 and Band 7, IAF, ALB, MSAVI2 and NDVI, see Table 2 for acronyms and abbreviations), which are correlated with many ecosystem attributes, such as photosynthetic activity, total plant cover, plant and soil moisture, plant stress and biomass (Lu *et al.,* 2004; Günlü *et al.,* 2014). Several studies have demonstrated that spectral bands and vegetation indices are usually good predictors for estimating AGB (Lu *et al.,* 2012; Castillo-Santiago *et al.,* 2013; Lu *et al.,* 2016; López-Serrano *et al.,* 2016a, 2016b, 2016c). Terrain features 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 (McNab 1989, Roberts and Cooper 1989, Wilson and Gallant 2000). Finally, the texture features may address some of the existing problems with vegetation index saturation and the data acquisition constraints related to mapping forest biomass at regional scales (Kelsey and Neff, 2014).

The results obtained in the statistical study of the 99 plots showed that the SVM technique yielded the best fits once the parameterization had been optimized (averaged ranking of 161.18, which is about 15% better than kNN, the second best technique), thus confirming that this type of technique is of great potential for improving biomass estimation, independently of the type of sensor to which it is applied, as demonstrated in recent studies (e.g. Zhao *et al.,* 2011; García-Gutiérrez *et al.,* 2015). However, the results show that SVMs are very sensitive to parameterization, which hampers their use by non-experts. For non-experts, an auto-parameterization procedure such as Grid Search, which is a classic technique used to fit machine learning models (Gleason and Im, 2012), could be applied. Unfortunately, this type of procedure has an important drawback as it separates optimization of parameters (specific to each technique) from feature selection. Both concepts (parameterization and feature selection) are closely related and should occur simultaneously (Huang and Wang, 2011). Nonetheless, Grid Search represents a simpler alternative to more complex procedures such as metaheuristics.

A boxplot with the AGB estimations obtained for the sample plots with the different approaches used after just one evolutionary parameterization and feature selection is shown in Figure 9. Both MLR and RF present a range of AGB estimations in the study area similar to the values observed in the sample plots used as training data (2 to 234 Mg ha^{-1}), especially MLR, although MLR tended to overestimate the values (Figure 9). However, the kNN and

SVM techniques estimated a limited range of values of AGB (from 56 to 138 Mg ha^{-1} for kNN and from 56 to 160 Mg ha^{-1} for SVM). This was mainly due to inaccurate parameterization by the evolutionary procedure (see averaged models vs. optimal models in Figures 6 and 7). In the case of kNN, inadequate feature selection may lead to a decrease in accuracy due to the Hughes effect. For the SVM, the number of parameters was higher and the evolutionary procedure was therefore more complex. Note that if the penalty factor (parameter C) is not well fitted and the hyperplane is thus not optimized, problems related to over or underfitting may occur (Xie *et al.,* 2008). In addition to the special random nature of evolutionary computation, this risk makes the automatic configuration for SVM difficult in a single optimization procedure (due to the random nature of evolutionary computation). Independently of whether automatic or a manual parameterization is selected, determination of the best configuration for remote sensing non-parametric techniques (especially for the most complex such as SVMs and not so much for others such as RF) is also time-consuming, scenario-dependent and sometimes requires a priori knowledge (Camps-Valls and Bruzzone, 2005).
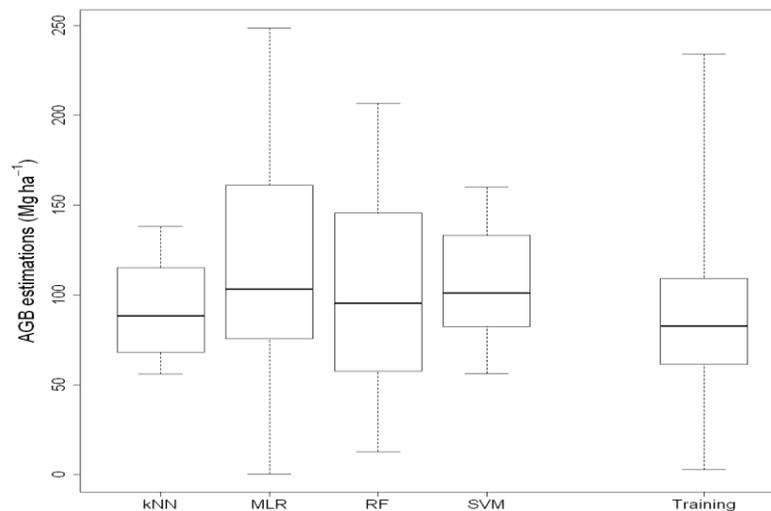
Figure 9. Box-plot of AGB estimations in the study area for the four techniques used (right) and AGB observed values in the sample plots (training data). Boxes represent the interquartile range, and maximum and minimum of AGB estimations are represented by upper and lower whiskers, respectively.

Conclusions

The kNN, RF and SVM machine learning algorithms are powerful tools for estimating aboveground forest biomass with remote sensing datasets, and they are all viable and accurate alternatives to the classic parametric MLR method. In addition to the usual sources of uncertainty associated with the accuracy of the AGB estimations from remote sensing data (field measurement errors, plot locations errors, errors of the individual tree biomass equations or error caused by geometrical and radiometric correction of remotely sensed data), parameterization of machine learning algorithms also has an important influence on the final performance of the models. The choice of method used will largely depend on the user's capacity to carry out that parameterization, because the techniques (especially SVMs) are not easy to apply and require a certain degree of expertise. Our findings indicate that SVM is the best alternative for experts, whereas RF represents a balance between model accuracy and ease of use for non-experts although differences with kNN could not be statistically demonstrated.

Acknowledgments

References

Abedi, M., Norouzi, G.H., Bahroudi, A., 2012. Support vector machine for multi-classification of mineral prospecting areas. Computers & Geosciences, Vol. 46, pp. 272-283.

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, Vol. 39, No. 6, pp. 507-520.

Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., Notarnicola, C., 2015. Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. Remote Sensing, Vol. 7, pp. 16398–16421.

Andersen, H.-E., Stunk, J., Temesgen, H., Atwood, D., Winterberger, K., 2011.Using multi-level remote sensing and ground data to estimate forest biomass resources in remote regions: A case study in the boreal forests of interior Alaska. Canadian Journal of Remote Sensing, Vol. 37, No. 6, pp. 596-611.

Asrar, G., 1989. Theory and applications of optical remote sensing, J. Wiley & Sons, New York. Pp 752.

Asrar, G., Fuchs, M., Kanemasu, E.T., Hatfield, J.L., 1984. Estimating absorbed photosynthetically active radiation and leaf area index from spectral reflectance in wheat. Agronomy Journal, Vol. 76, No. 2, pp. 300-306.

Baret, F., Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sensing of Environment, Vol. 35, No. 23, pp. 161-173.

Bayoudh, M., Roux, E., Richard, G., Nock, R., 2015. Structural knowledge learning from maps for supervised land cover/use classification: Application to the monitoring of land cover/use maps in French Guiana. Computers & Geosciences, Vol. 76, pp. 31-40.

Breidenbach, J., Næsset, E., Gobakken, T., 2012. Improving k-nearest Neighbor Predictions in Forest Inventories by Combining High and Low Density Airborne Laser Scanning Data. Remote Sensing of Environment, Vol. 117, pp. 358–365.

Breiman, L., 2001. Random Forests. Machine Learning, Vol. 45, No. 1, pp. 5–32.

Brutsaert, W., 1975. On a derivable formula for long-wave radiation from clear skies. Water Resources Research, Vol. 11, No. 5, pp. 742-744.

Bunker, D.E., DeClerck, F., Bradford, J.C., Colwell, R.K., Perfecto, I., 2005. Species loss and aboveground carbon storage in a tropical forest. Science, Vol. 310, pp. 1029–1031.

Camps-Valls, G., Bruzzone, L. 2005. Kernel-based methods for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, Vol. 43, pp. 1351−1362.

Castillo-Santiago, M.A, Ghilardi, A., Oyama, K., Hermadez-Stefanoni, J.L., Torres, I., Flamenco-Sandoval, A., Fernández, A., Jean-Francois, M., 2013. Estimating the spatial distribution of Woody biomass suitable for charcoal making from remote sensing and geostatistics in central Mexico. Energy for Sustainable Development, Vol. 17, No. 2, pp. 177-188.

Corral-Rivas, J., Vargas, B., Wehenkel, C., Aguirre, O., Álvarez, J.G., Rojo, A., 2009. Guía para el Establecimiento de Sitios de Inventario Periódico Forestal y de Suelos del Estado de Durango. Universidad Juárez del Estado de Durango. pp 81.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning, Vol. 20, No. 3, pp. 273-297.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, Vol. 13, No. 1, pp. 21-27.

Cracknell, J.M., Reading, M.A., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Computers & Geosciences, Vol. 63, pp. 22-33.

Deschamps, B., McNairn, H., Shang, J., Jiao, X, 2012. Towards operational radar-only crop type classification: comparison of a traditional decision tree with a random forest classifier. Canadian Journal of Remote Sensing, Vol. 38, No. 1, pp. 60-68.

Dube, T., Mutanga, O., 2015. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in Mgeni catchment, South Africa. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 101, pp. 36–46.

Eckert, S., 2012. Improved forest biomass and carbon estimations using texture measures from worldview-2 satellite data. Remote Sensing, Vol. 4, No. 4, pp. 810-829.

ERDAS Inc., 2014. Erdas Imagine. Available online:

http://www.hexagongeospatial.com/products/ERDAS-IMAGINE/details.aspx (accessed on 10/06/2014)

FAO, 2001. Global Forest Resources Assessment 2000, Main Report. FAO Forestry Paper 140. Rome: FAO.

FAO, 2006. Global Forest Resources Assessment 2005. FAO Forestry Paper 147. Rome: FAO.

Fassnacht, F.E., 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, Vol. 154, pp. 102-114.

Frazier, R. J., Coops, C.N., Wulder, A.M., Kennedy, R., 2014. Characterization of aboveground biomass in an unmanaged boreal forest using Landsat temporal segmentation metrics. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 92, pp. 137-146.

García, S., Fernández, A., Luengo, J., Herrera, F., 2010. Advanced non-parametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. Information Sciences, Vol. 180, No. 10, pp. 2044-2064.

García-Gutiérrez, J., González-Ferreiro, E., Riquelme-Santos, J.C., Miranda, D., Dieguez-Aranda, U., Navarro-Cerrillo, R. M., 2014. Evolutionary feature selection to estimate forest stand variables using LiDAR. International Journal of Applied Earth Observation and Geoinformation, Vol. 26, pp. 119-131.

García-Gutiérrez, J., Martinez-Alvarez, F., Troncoso, A., Riquelme, J. C. 2015. A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. Neurocomputing, Vol. 167, pp. 24-31.

GCOS, 2010. Global Climate Observing System. Essential Climate Variables. Available online: http://www.wmo.int/pages/prog/gcos/index.php?name=Essential ClimateVariables. (Accessed on 13/04/2016).

Geosystems GmbH., 2013. Atcor for Erdas Imagine. Geosystems GmbH, Switzerland.

Ghosh, A., Fassnacht, F.E., Joshi, P.K., Koch, B., 2014. A framework for map ping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. International Journal of Applied Earth Observation and Geoinformation, Vol. 26, pp. 49-63.

Günlü, A., Ercanli, A., Baskent, I., Cakir, G. 2014. Estimating aboveground biomass using Landsat TM imagery: A case study of Anatolian Crimean pine forests in Turkey. Annals of Forest Research, Vol. 57, pp. 289-298.

Guo, Y., Tianb,X., Lib,Z., Linga,F., Chenb,E., Yan, M., Lib, C., 2014. Comparison of estimating forest above-ground biomass over montane area by two non-parametric methods. In: Proceedings of 2014 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 741-744.

Hall, M., 1999. Correlation-based Feature Selection for Machine Learning. PhD thesis, University of Waikato, Hamilton, New Zealand. pp. 178.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H., 2009. The WEKA data mining software: An update. SIGKDD Explorations, Vol. 11, No. 1, pp. 10-18.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics, Vol. 3, No. 6, pp. 610-621.

Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O'Hara, R.B., Scheiter, S., Huth, A., 2012. Connecting dynamic vegetation models to data - an inverse perspective. Journal of Biogeography, Vol. 39, No. 12, pp. 2240-2252.

Huang, C.L., Wang, C.J., 2006. A GA-based feature selection and parameters optimization for support vector machines. Expert Systems with Applications, Vol. 31, No. 2, pp. 231-240.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, Vol. 25, No. 3, pp. 295-309.

Hughes, G.F. 1968. On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory, Vol. 14, No. 1, pp. 55–63.

INEGI - Instituto Nacional de Estadística Geográfica e Informática, 2012. Uso del suelo y vegetación escala 1: 250 000 serie v, información vectorial. México.

INEGI - Instituto Nacional de Estadística Geográfica e Informática, 2014. Continuo de elevaciones mexicano 3.0 (CEM 3.0). México.

Jachowski, N.R.A., Quak, M.S.Y., Friess, D.A., Duangnamon, D., Webb, E.L., Ziegler, A.D. 2013. Mangrove biomass estimation in Southwest Thailand using machine learning. Applied Geography, Vol. 45, pp. 311–321.

Jung, J., Kim, S., Hong, S., Kim, K., Kim, E., Im, J., Heo. J., 2013. Effects of National Forest Inventory Plot Location Error on Forest Carbon Stock Estimation Using k-nearest Neighbor Algorithm. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 81, pp. 82–92.

Kelsey, K.C., Neff, J.C., 2014. Estimates of Aboveground Biomass from Texture Analysis of Landsat Imagery. Remote Sensing, Vol. 6, No. 7, pp. 6407-6422.

Latifi, H., Nothdurft, A., Koch, B., 2010. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/lidar-derived predictors. Forestry, Vol. 83, No. 4, pp. 395-407.

Laurin, V.G., Chen, Q. Lindsell, A.J., Coomes, A.D., Del Frate, F., Guerriero, L., Pirotti, F., Valentini, R., 2014. Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 89, pp. 49-58.

Le Toan, T., Quegan, S., 2015. BIOMASS Biomass Monitoring Mission for Carbon Assessment. CESBIO. Toulouse, France.

López-Serrano, P.M., Corral-Rivas, J.J., Díaz-Varela, R.A., Álvarez-González, J.G., López-Sánchez, C.A., 2016a. Evaluation of Radiometric and Atmospheric Correction Algorithms for Aboveground Forest Biomass Estimation Using Landsat 5 TM Data. Remote Sensing, Vol. 8, 369.

López-Serrano, P.M., López-Sánchez, C.A., Díaz-Varela, R.A., Corral-Rivas, J.J., Solís-Moreno, R., Vargas-Larreta, B., Álvarez-González, J.G., 2016b. Estimating biomass of mixed and uneven-aged forests using spectral data and a hybrid model combining regression trees and linear models. iForest - Biogeosciences and Forestry, Vol. 9 pp. 226-234.

López-Serrano, P.M., López-Sánchez, C.A., Solís-Moreno, R., Corral-Rivas, J.J., 2016c. Geospatial Estimation of above Ground Forest Biomass in the Sierra Madre Occidental in the State of Durango, Mexico. Forests, Vol. 7, 70.

Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing, Vol. 27, pp. 1297–1328.

Lu, D., Mausel, P., Brondízio, E., Moran, E. 2004. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. Forest Ecology and Management, Vol. 198, pp. 149-167.

Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., Moran, E., 2016. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. International Journal of Digital Earth, Vol 9, No. 1, pp 63-105.

Lu, D., Qi, C., Guangxing, W., Moran, E., Batistella, M., Maozhen Z., Vaglio, L.G., Saah, D., 2012. Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. International Journal of Forestry Research. Vol. 2012, pp. 1-16.

Luengo, J., García, S., Herrera, F., 2009. A study on the use of statistical tests for experimentation with neural networks: Analysis of parametric test conditions and non-parametric tests. Expert Systems with Applications, Vol. 36, No. 4, pp. 7798-7808.

Malhi, Y., Meir, P., Brown, S., 2002. Forests, carbon and global climate. Philosophical Transactions of the Royal Society A, Vol. 360, pp. 1567–1591.

Malhi, Y., Phillips, O.L., 2004. Tropical forests and global atmospheric change: a synthesis. Philosophical Transactions of the Royal Society, Vol. 359, pp. 549–555.

Mateos, G. J., García, G., Riquelme, J.C.S., 2012. On the evolutionary optimization of k-NN by label-dependent feature weighting. Pattern Recognition Letters, Vol. 33, No. 16, pp. 2232-2238.

McNab, W.H., 1989. Terrain shape index: quantifying effect of minor landforms on tree height. Forest Science, Vol. 35, pp. 91-104.

McRoberts, R.E., 2012. Estimating forest attribute parameters for small areas using nearest neighbors techniques. Forest Ecology and Management, Vol. 272, pp. 3-12.

Molinaro, A. M., Simon, R., Pfeer, R. M., 2005. Prediction error estimation: a comparison of resampling methods. Bioinformatics, Vol. 21, No. 15, pp. 3301-3307.

Moore, I.D., Nieber, J.L., 1989. Landscape assessment of soil erosion and non-point source pollution. Journal of the Minnesota Academy of Science, Vol. 55, pp. 18-24.

Mutanga, O., Skidmore, A.K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. International Journal of Remote Sensing, Vol.25, pp. 3999–4014.

Mutanga, O., Adam, E., Cho, M.A., 2012. High Density Biomass Estimation for Wetland Vegetation Using WorldView-2 Imagery and Random Forest Regression Algorithm. International Journal of Applied Earth Observation and Geoinformation, Vol. 18, pp. 399–406.

Næsset, E., Gobakken, T., Bollandsås, O.M., Gregoire, T.G., Nelson, R., Ståhl. G., 2013. Comparison of Precision of Biomass Estimates in Regional Field Sample Surveys and Airborne LiDAR-assisted Surveys in Hedmark County, Norway. Remote Sensing of Environment, Vol. 130, pp. 108–120.

Packalén, P., Temesgen, H., Maltamo, M. 2012. Variable selection strategies for nearest neighbor imputation methods used in remote sensing based forest inventory. Canadian Journal of Remote Sensing, Vol. 38, No. 6, pp. 557-569.

PCI Geomatics Inc., 2013. PCI Geomatics.

Pflugmacher, D., Cohen, W.B., Kennedy, R.E., Yang, Z., 2014. Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics. Remote Sensing of Environment, Vol. 151, pp. 124–137.

Pretzsch, H., Biber, P., Ďurský, J., Gadow, K.v., Hasenauer, H., Kändler G., Kenk, G., Kublin, E., Nagel, J., Pukkala, T., Skovsgaard, J.P., Sodtke, R., Sterba, H., 2002. Recommendations for standardized documentation and further development of forest growth simulators. Forstwissenschaftliches Centralblatt. Vol. 121, No. 3, pp. 138-151.

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., 1994. A modified soil adjusted vegetation index. Remote Sensing of Environment, Vol. 48, No. 2, pp. 119-126.

R Core Team, 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, pp. 3551.

Roberts, D.W., Cooper, S.V., 1989. Concepts and techniques of vegetation mapping. In Land Classifications Based on Vegetation: Applications for Resource Management. USDA Forest Service GTR INT-257. Ogden, UT. pp. 90-96.

Rouse, J.W., Haas, R.H., Schiell, J.A., Deferino, D.W., Harlan, J.C., 1974. Monitoring the vernal advancement of retrogradation of natural vegetation. NASA/OSFC Type III, Oreenbello MD. pp. 112.

Samadzadegan, F., Hasani, H., Schenk, T., 2012. Simultaneous feature selection and SVM parameter determination in classification of hyperspectral imagery using Ant Colony Optimization, Canadian Journal of Remote Sensing, Vol. 38, No. 2, pp. 139-156.

Shataee, S., 2013. Forest Attributes Estimation Using Aerial Laser Scanner and TM Data. Forest Systems, Vol. 22, No. 3, 484-496.

Tian, X., Li, Z., Su, Z., Chen, E., Van der Tol, C., Li, X., Guo,Y., Li L., Ling, F., 2014. Estimating montane forest above-ground biomass in the upper reaches of the Heihe River Basin using Landsat-TM data. International Journal of Remote Sensing, Vol. 35, No. 21, pp. 7339-7362.

Tian, X., Su, Z., Chen, E., Li, Z., van der Tol, C., Guo, J., He, Q., 2012. Estimation of Forest Above-ground Biomass Using Multi-parameter Remote Sensing Data Over a Cold and Arid Area. International Journal of Applied Earth Observation and Geoinformation, Vol. 14, No. 1, pp. 160–168.

Vargas-Larreta, B., 2013. Estimación del potencial de los bosques de Durango para la mitigación del cambio climático. Modelización de la biomasa forestal. Proyecto: FOMIX-DGO-2011-C01-165681. ITES, EL Salto (México).

Wang, L., Zhou, X., Zhu, X., Dong, X., Guo, W., 2016. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data, The Crop Journal, *in press*.

Wehenkel, C., Corral-Rivas, J.J., Hernández-Díaz, J.C., Gadow, K.v., 2011. Estimating Balanced Structure Areas in multi-species forests on the Sierra Madre Occidental, Mexico. Annals of Forest Science, Vol. 68, pp. 385-394

Wilson, J.P., Gallant, J.C., 2000. Digital terrain analysis. In: Terrain Analysis: Principles and Applications. Edited by JohnWiley & Sons, New York. pp 1-28.

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, Vol. 8 No. 2, 159.

Xie, X., Liu, W.T., Tang, B. 2008. Space-based estimation of moisture transport in marine atmosphere using support vector regression, Remote Sensing of Environment. Vol. 112, No. 4, pp. 1846-1855.

Yang, Y., Monserud, R.A., Huang, S., 2004. An evaluation of diagnostic tests and their roles in validating forest biometrics models. Canadian Journal of Forest Research. Vol. 34, pp. 619-629.

Zhang, Y., Liang, S., Sun, G., 2014. Forest biomass mapping of northeastern China using GLAS and MODIS data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol, 7, pp. 140–152.

Zhao, K., Popescu, S., Meng, X., Pang, Y, Agca, M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sensing of Environment. Vol. 115, No. 8, pp. 1978-1996.

Zhu, X., Liu, D., 2015. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 102, pp. 222–231.

CAPÍTULO 5. EVALUATION OF RADIOMETRIC AND ATMOSPHERIC CORRECTION ALGORITHMS FOR ABOVEGROUND FOREST BIOMASS ESTIMATION USING LANDSAT-5 TM DATA

Pablito M. López-Serrano[1], José J. Corral-Rivas[2], Ramón A. Díaz-Varela[3], Juan G. Álvarez-González[4], Carlos A. López-Sánchez[2]*

Abstract

Solar radiation is affected by absorption and emission phenomena during its downward trajectory from the sun to the earth's surface and during the upward trajectory detected by satellite sensors. This leads to distortion of the ground radiometric properties (reflectance) recorded by satellite images, used in this study to estimate aboveground forest biomass (AGB). Atmospherically-corrected remote sensing data can be used to estimate AGB on a global scale and with moderate effort. The objective of this study was to evaluate four atmospheric correction algorithms (for surface reflectance) (ATCOR2, FLAASH, COST and 6S) and one radiometric correction algorithm (for reflectance at the sensor) (ToA) to estimate AGB in temperate forest in the northeast of the state of Durango, Mexico. The AGB was estimated from Landsat-5 TM imagery and ancillary information from a digital elevation model (DEM) using non parametric Multivariate Adaptive Regression Splines (MARS) technique. Field reference data for the model training were collected by systematic sampling of 99 permanent monitoring sites established during the winter of 2011. The following predictor variables were identified in the MARS model: *Band7*, *Band5*, *slope* (*β*), *NDVI* and *MSAVI2*. After cross validation, 6S was found to be the optimal model for estimating AGB ($R^2 = 0.71$ and RMSE=33.5 Mg ha^{-1}, 37.61% of the average stand biomass). We conclude that atmospheric and radiometric correction of satellite images can be used along with non-parametric techniques to estimate AGB with acceptable accuracy.

Keywords: Multivariate Adaptive Regression Splines, Remote Sensing, Radiometric Correction Algorithms, Terrain Features.

[1]Doctorado en Ciencias Agropecuarias y Forestales, DICAF. Universidad Juárez del Estado de Durango.
[2]Instituto de Silvicultura e Industria de la Madera. Universidad Juárez del Estado de Durango.
[3]Departamento de Botánica – GI-1908-BIOAPLIC. Universidad de Santiago de Compostela. Escuela Politécnica Superior, Lugo, España.
[4]Departamento de Ingeniería Agroforestal. Universidad de Santiago de Compostela. Escuela Politécnica Superior, Lugo, España.
*Autor para correspondencia: calopez@ujed.mx

Introduction

Apart from sensor gains/offsets, solar irradiance and Sun-Earth geometry, the characteristics of electromagnetic energy detected by remote sensing optical sensors is affected by particles and other components present in the atmosphere. Hence, as solar radiation passes through the atmosphere (Sun-surface-sensor), it is affected by absorption or scattering by particles in suspension (aerosols) and other atmospheric elements, thus creating a hazy effect that distorts the radiometric properties of satellite images (Jensen, 1996; Gao *et al*., 2009; Tyagi and Bhosle, 2011; Tan *et al*., 2012). Various algorithms have been developed to correct such effects in the image. These include calibrations based on sensor parameters, solar-Earth geometry, dark object subtraction and radiative transfer (Hadjimitsis *et al*., 2010; Li *et al*., 2012; Jaelani *et al*., 2015; Lee *et al*., 2015). Correction of atmospheric effects is important in relation to improving data quality (Vanonckelen *et al*., 2013; Moreira *et al*., 2014). However, the use of correction algorithms to minimize atmospheric effects, especially the scattering and absorption caused by aerosols, remains challenging (Roy *et al*., 2014). This particularly applies to the parameterization of algorithms for calculating the surface reflectance for its eventual use in estimating aerial or aboveground forest biomass (AGB) (Ju *et al*., 2012).

AGB is an important parameter that provides information about the current status of forests and therefore facilitates forest management decisions (Ediriweera *et al*., 2013; Gagliasso *et al*., 2014). The use of optical remote sensors of medium spatial resolution, such as the Landsat Thematic Mapper (TM), has become increasingly common in the last few decades. The use of these sensors for large-scale monitoring of AGB dynamics, particularly in the case of multi-temporal/multi-scene analyses or unevenly-distributed atmospheric effects in single-image analyses, frequently involves the application of different atmospheric correction algorithms. Such algorithms correct distortions between or within images other than those related to real land cover differences (Hantson and Chuvieco, 2011; Balthazar *et al*., 2012; Cortés *et al*., 2014; Kelsy and Neff, 2014; Tian *et al*., 2014).

The aim of the present study was to evaluate four different atmospheric correction algorithms (surface reflectance)—COST (Cosine of the Solar Zenith Angle), a modification of the dark object subtraction method (Vicent, 1972) with the addition of a multiplicative correction for atmospheric transmittance (Chavez, 1996), ATCOR2 (Atmospheric Correction for Flat Terrain), (Richter, 1996), FLAASH (Fast Line-of-sight Atmospheric Analysis of

Spectral Hypercubes, (Adler *et al*., 1998)) and 6S (Second Simulation of Satellite Signal in the Solar Spectrum, (Vermote *et al*., 1997), and one radiometric correction algorithm (apparent reflectance at sensor), ToA (Top of Atmosphere), for estimating AGB from spectral data captured by the Landsat 5 TM sensor and topographic variables derived from a digital elevation model (DEM) generated using data collected by sampling of the permanent forest growth and soil research sites (SPIFyS) established in a temperate forest in the northwest of the state of Durango, Mexico.

Different methods could be used for remote AGB estimation in forests, including parametric and nonparametric approaches. However, since forest structure and biomass often entail nonlinear variability, variable interaction across scales and autocorrelation, nonparametric approaches often markedly outperform parametric methods (Saatchi *et al*., 2001).

In this study, we investigate remote AGB estimation of mixed and uneven-aged forests using the multivariate adaptive regression splines (MARS). To the best of our knowledge, MARS have rarely been used for remote estimation of AGB (Moisen and Frescino, 2002; Güneralp *et al*., 2014, Filippi *et al*., 2014), but in these cases, the method has performed best for prediction than other parametric and non-parametric approaches, such us linear models, classification, regression trees and artificial neural networks (Moisen and Frescino, 2002) or hybrid tree-based algorithms (Moisen and Frescino, 2002; Güneralp *et al*., 2014).

Materials and Methods

Study Area

The study area is located in the northwest of the state of Durango (Figure 1). The forest vegetation comprises pine (*Pinus*spp.), oak (*Quercus* spp.), fir (*Abies* spp.) and plurispecific stands with different proportions of species of the genera *Pinus*and *Quercus*, in accordance with the Land and Vegetation Use Map, scale 1:250,000, Series V [30]. The weather is cold in the canyons (with temperatures ranging from −20 °C in winter to 20 °C in summer) and mildor warm in the valleys (with temperatures ranging from 10 °C in winter to 40 °C in summer).

Field Data

The field data were AGB values taken from a database of stand variables measured in 99 permanent forest growth and soil research sitesSPIFyS established during the winter of 2011 in accordance with the method developed by Corral-Rivas *et al.* (2009). The SPIFyS are square plots of a size of 50 × 50 m (surface area, 2500 m^2). The location of the sites was carried out by systematic sampling at the regional level using a squared grid of five kilometers, although a small squared grid (3 km) was used on very steep areas to include their variability.
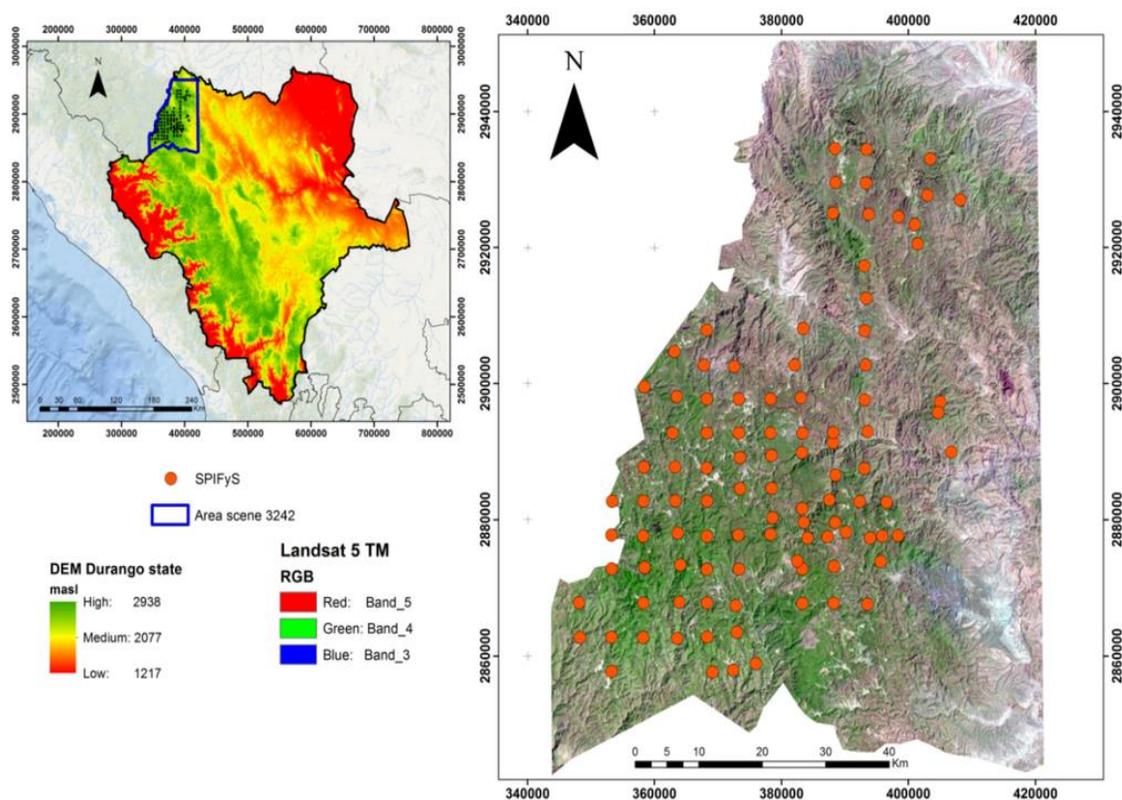
Figure 1. Location of the study area and of the permanent forest growth and soil research sites (SPIFyS).

In each sample plot, the tree species were recorded, and the diameter at breast height (cm) and total height (m) of all standing trees were measured. Species-specific models developed for the study area were used to estimate individual tree aboveground biomass (Vargas-Lagarreta, 2013). The goodness of fit for such statistical models ranged from 0.87–0.99 (R^2),

and the root mean square error (*RMSE*) varied from 22.8–95.2 kg. Once the tree aboveground biomass was estimated, all values from each sampling plot were summed and expressed on a per hectare basis. Table 1 shows the descriptive statistics of the main stand variables.

Spectral Data from the Landsat 5 TM Sensor

The image used in the study, which was captured by the Landsat 5 TM sensor in April 2011, is available at the U.S. Geological Service (USGS) website [33]. The sensor operates in seven bands of the electromagnetic spectrum: *Band 1* (blue, 450–520 nm), *Band 2* (green, 520–600 nm), *Band 3* (red, 630–690 nm), *Band 4* (near-infrared, 760–900 nm), *Bands 5* and *7* (mid-infrared region, 1550–2350 nm) and *Band 6* (which provides information in the thermal infrared region and is not used in this type of study). Likewise, the spectral indices Normalized Difference Vegetation Index (*NDVI*, (Rouse *et al.*, 1974)) and Modified Soil Adjusted Index (*MSAVI2*, (Baret and Guyot, 1991)) were used. Such indexes are potentially less sensitive than single band values to artefacts due to differences in light conditions, exposed soil or terrain slope and orientation, which might eventually affect AGB estimation.

Table 1. Descriptive statistics of the main stand variables estimated from the 99 permanent sample plots. The dominant stand height was calculated as an average value from the 100 thickest trees per hectare.

Variable	Mean	Standard deviation	Minimum value	Maximum value
Number of stems per ha	655.47	322.25	224	2264
Diameter at breast height per ha (cm)	18.44	3.46	11.69	31.12
Dominant height (m)	14.62	3.72	6.87	24.81
Stand biomass (Mg ha^{-1})	89.03	43.45	2.70	234.03

2.4. Radiometric Correction Algorithms

A useful prior step to the interpretation of satellite images is converting the digital numbers (DNs) stored in the original image into biophysical variables of standard significance (reflectance). These variables are comparable in the same sensor over time and over scenes, as well as between different sensors and between remote sensing and other methods of detecting electromagnetic energy. The correction is also advisable in the case of unevenly-distributed atmospheric effects in the image and also when vegetation indexes based on band ratios are

included in the analyses (Lavreau, 1991; Guyot and Gut, 1994; Richter, 1996; Carlotto, 1999; Carlotto, 1999; Song *et al.*, 2001; Liang *et al.*, 2002). In order to calculate surface reflectance, the atmospheric effects must be removed. This involves estimating the atmospheric transmissivity (both up and down welling), diffuse irradiance and atmospheric radiance due to scattering (Chuvieco, 2010). In the present study, we evaluated five different radiometric correction methods: the first four methods (namely ATCOR2, COST, FLAASH and 6S) including atmospheric corrections (surface absolute reflectance), while the fifth (ToA) without considering atmospheric effects (apparent reflectance at the sensor) for the final estimation of AGB.

Atmospheric Correction for Flat Terrain (ATCOR2)

The aim of ATCOR2 correction (Richter, 1996) is to remove the atmospheric effects in order to recover the physical parameters of the Earth's surface, including the surface reflectance, soil visibility and the temperature. The correction is carried out with the ATCOR module (Geosystems, 2013) included in ERDAS IMAGINE software, Version 2013(ERDAS, 2014).

Cosine of the Sun Zenith Angle (COST)

This is a radiometric calibration method that considers the atmospheric effect and is based entirely on the characteristics of the satellite image, in contrast with other methods of atmospheric correction, like ATCOR2, FLAASH or 6S, requiring some extra parameters, such as atmospheric profiles, the aerosol models or visibility (Chavez, 1992; 1996). COST applies Dark Object Subtraction (DOS, (Vicent, 1972)) to compensate for the additive components of the atmosphere, which mainly affect the shortest wavelengths. DOS does not take into account the multiplicative effect on the longer wavelengths. For initial estimation of the multiplicative effect, the value of the atmospheric transmittance along the path from the ground to the sensor (TAUz) was computed from the cosine of the solar zenith angle (Moran *et al.*, 2008). This correction was carried out by implementing the algorithm in the Model Maker$^{®}$ module of ERDAS IMAGINE software, Version 2013 (ERDAS, 2014).

Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH)

This is an advanced atmospheric correction module based on the MODTRAN4 algorithm of radiative transfer developed by Spectral Sciences Inc. (Burlington, MA, USA) under the

sponsorship of the U.S. Air Force Research Laboratory (Adler-Golden *et al.*, 1998; Anderson *et al.*, 2002). The technique is initially based on the standard equation of spectral radiance for each pixel of the sensor, which applies to the range of wavelengths of solar radiation (thermal emission is omitted) and a Lambertian and flat surface or their equivalents. It considers the radiance reflected from the Earth's surface and scattered by the atmosphere towards the sensor. The difference between these two radiances is due to the adjacency effect (spatial mixture of radiance among nearby pixels) caused by atmospheric scattering. The correction was carried out with the FLAASH module in ENVI® software, Version 5.1 (EXCELIS, 2013).

Second Simulation of Satellite Signal in the Solar Spectrum (6S)

This procedure eliminates atmospheric effects on the reflectance values in images captured by sensors on board satellite or aircraft platforms. It is based on an advanced code of radiative transfer, designed for simulating the interaction between solar radiation and an atmospheric-surface system, together with a wide range of atmospheric, spectral and geometric conditions (Vermote *et al.*, 1997). The code acts on the basis of the successive order of scattering (SOS) method and explains the polarization of radiation in the atmosphere by calculating the different components. The scenes used to belong to the National Landsat Archive Processing System (NLAPS) and correspond to the product obtained by Landsat 4–5 Thematic Mapper Level 1 of surface reflectance processed by the Standard Landsat Product Generation System (LPGS) and using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (available at the U.S. Geological Service website (USGS, 2011).

Apparent Reflectance at the Top of Atmosphere (ToA)

This technique enables calculation of the apparent reflectance in a satellite image and consists of converting the DNs to radiance values and then to reflectance values. The word "apparent" refers to the fact that the reflectance has not been corrected for atmospheric effects and represents an initial normalization of the image (Huang *et al.*, 2014). The correction was carried out with the ''apparent reflectance'' method implemented in the "landsat" package (Goslee, 2015) in the R software (R Core Team, 2015).

Parameters Derived from the Digital Elevation Model (DEM)

The primary and secondary terrain parameters were calculated from the DEM (Table 2), after the application of a low bandpass filter with a moving window of 5 × 5, with the aim of

reducing the banding effect present in the original archive (INEGI, 2014). These parameters are considered as predictor variables for estimating AGB, because they are directly related to the distribution and development of forest species and can subsequently be evaluated (MaNab, 1989; Roberts and Cooper, 1989).

Generation of a Database

The database used to estimate AGB was generated by extracting the mean values of the satellite image pixels after each of the radiometric correction techniques and of the different DEM parameters by considering a buffer of 25 m for the geolocalization of the SPIFyS plots. The values were extracted using the "raster" package (Hijmans, 2015), implemented in R statistical software (R Core Team, 2014).

Table 2. Additional field variables used to model the aboveground forest biomass (AGB).

Variable	Formula	Reference
Elevation	Digital Elevation Model	
Slope (β)	$\beta= \arctan\left[(G^2 + H^2)^{1/2}\right]$	
Transformed Aspect (*Trasp*)	$Trasp= \dfrac{1\text{-}\cos\left((\pi\,/\,180)(\alpha\text{-}\,30)\right)}{2}$	[42]
Terrain Shape Index (*TSI3*)	$TSI=\bar{Z}/R$	[41]
Wetness Index (*WI*)	$WI= \ln (As/\tan\beta)$	[43]
Profile curvature (Ø)	$\emptyset=-2\dfrac{DG^2+EH^2+FGH}{G^2+H^2}$	
Plan curvature (ω)	$\omega=2\dfrac{DH^2+EG^2+FGH}{G^2+H^2}$	[44]
Curvature *(x)*	$x= \omega - \emptyset$	

where:
$\bar{Z}:$ Average elevation.
$R:$ Point radio altitude units.
$As:$ Drainage area specified.
$\tan\beta:$ Local slope angle.
D, F, G and H were derived according to equation of Chavez (1992).

Statistical Analysis

Analysis of Variance

In order to establish whether there were any significant differences between the radiometric correction algorithms considered, analysis of variance (ANOVA) was used to compare the values of each spectral band of the sensor obtained by processing the images of the SPIFyS and applying the different correction algorithms. The correction algorithm was used as a fixed factor:

$$yi = \mu + CA_i + \varepsilon_i \tag{1}$$

where y_i are the values of each spectral band of the sensor corrected by the algorithm i, CA_i is the correction algorithm used (fixed factor) and ε_i are the associated errors.

Prior to the application of the ANOVA, the dependent variables normality and homogeneity of variance were checked using respectively Shapiro–Wilk's test (Shapiro and Wilk, 1965) and Levene's test (Levene, 1960). When the results of these tests indicated that the ANOVA assumptions were not satisfied, the dependent variables were transformed by a natural log transformation. When the ANOVA indicated significant differences between correction algorithms ($\alpha = 0.05$), Tukey's multiple comparison test was used to identify which algorithms were different. Likewise, box plots were elaborated for each variable (spectral band) with the aim of facilitating graphical interpretation of the results. All analyses were carried out with R statistical software (R Core Team, 2014).

Multivariate Adaptive Regression Spline (MARS)

The statistical analysis was carried out by the non-parametric multivariate adaptive regression splines (MARS) proposed by Friedman (Friedman, 1991), using the ''earth'' package (Molborrow, 2015) implemented in the R software (R Core Team, 2014). This method involves constructing a non-linear regression model based on a product of functions denominated ''smoothed basis functions'' (splines). The predictors are incorporated in their structure as part of a function generating a model for the dependent variable, which may be continuous or binary, and at the same time automatically selects the predictors in the final model. The general form of the MARS non-parametric regression model, formulated on the dependent ''y'' variable and the ''x'' predictors, is as follows:

$$y = f(x) + \varepsilon \tag{2}$$

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

$$f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(x) \tag{3}$$

where β_0 is the intercept of the model, $B_m(x)$ is the basis function of the mth base, βm is the coefficient of the mth base, and M is the number of bas is 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 a constant value, ii) a product of two or more hinge functions, that, therefore, can model interaction between two or more predictors (x).

The optimal model was selected using the backward procedure: an overfitted model was generated with all possible predictor variables and the model was pruned by removing terms one by one, deleting the least effective term at each step until the best submodel was found. Model subsets were compared using the generalized crossed validation (GCV) criterion. GCV is an approximation to the error that would be determined by leave-one-out validation and considers both the residual error and the model complexity, evaluated, the optimal model will therefore be that yielding the lowest GCV.

$$GCV(M) = \frac{\sum_{i=1}^{n}(y_i - f_M(x_i))^2}{n\left(1 - \frac{p_M}{n}\right)^2} \tag{4}$$

where y_i are 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 are the observed values of the predictors included in the MARS model and p_M is the number of parameters of the MARS model.

To analyse the importance of the independent variables that contribute most to the final model, three different criteria were used: i) *nsubsets*, which represents the number of times

that each variable is included in a subset (in the final model), ii) *sqr rss*, which first calculates the decrease in the sum of square errors (*RSS*) for each subset relative to the previous subset, and iii) *sqr gcv,* which uses the same calculation process, but with *GCV* instead of *RSS*. The criteria *sqr rss* and *sqr gcv* are normalized on a scale of 100, to facilitate interpretation of the contribution of each predictive variable in the model.

With the aim of evaluating the performance of the fitted model for each of the radiometric correction algorithms, the following statistics were determined to compare the AGB reference data for the monitoring plots with the estimates based on the images subjected to different corrections: root mean squared error (*RMSE*), the coefficient of determination (R^2) and the generalized coefficient of determination (GR^2):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - p}} \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

$$GR^2 = \frac{1 - GCV}{GCV_{null}} \tag{7}$$

where y_i, $\hat{y}_i$ and $\bar{y}$ are respectively the observed, estimated and mean values of the dependent variable, n is the total number of observations used to fit the model, p is the number of parameters to estimate, and GCV_{null} is the *GCV* of a model with a single independent term (β_0).

Because the quality of the fit does not necessarily reflect the quality of the prediction, assessment of the validity of the model with an independent dataset is desirable (Myers, 1990). Therefore, cross-validation was carried out by splitting the dataset into different numbers of folds (2–10), and the goodness-of-fit statistics (Equations (5)–(7)) were calculated for each fold (test data) by using the models fitted to the other folds (training data).

Generation of Thematic Maps

The final stage of the overall process illustrated was the calculation of AGB maps. These maps were derived by implementing the basic functions resulting from the MARS model fit

for each of the atmospheric correction algorithms by using the raster calculator in ArcGIS® software (ESRI, 2011). An overview of the workflow described in the previous sections is presented in Figure 2.
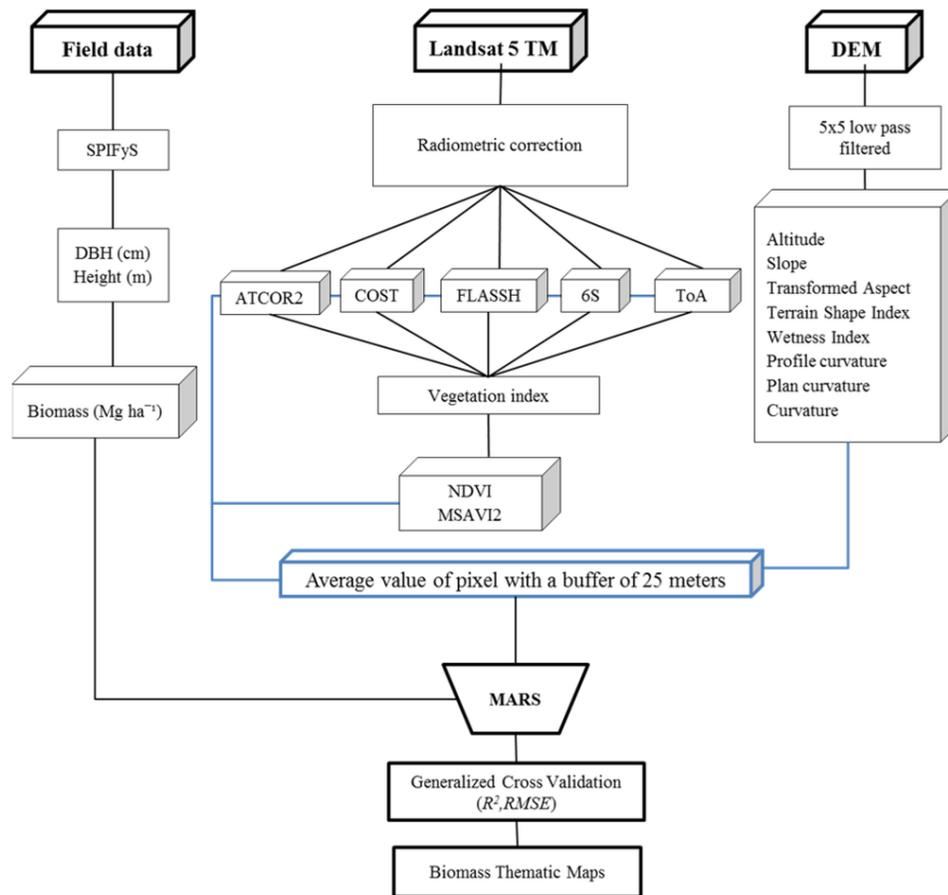
Figure 2. Flow diagram of the processes involved in estimating AGB.

Results

In the first step, a visual comparison of the application of the different radiometric correction algorithms to the Landsat 5 TM image of the study area was done. All atmospheric correction algorithms were parameterized using the same type of "rural" aerosol, for zones not affected by urban or industrial sources and a visibility larger than 40 km. In comparison with the original image in DNs, the brightness of the other images differs due to the parameterization used in each algorithm to correct the radiometric and atmospheric effects (Figure 3).
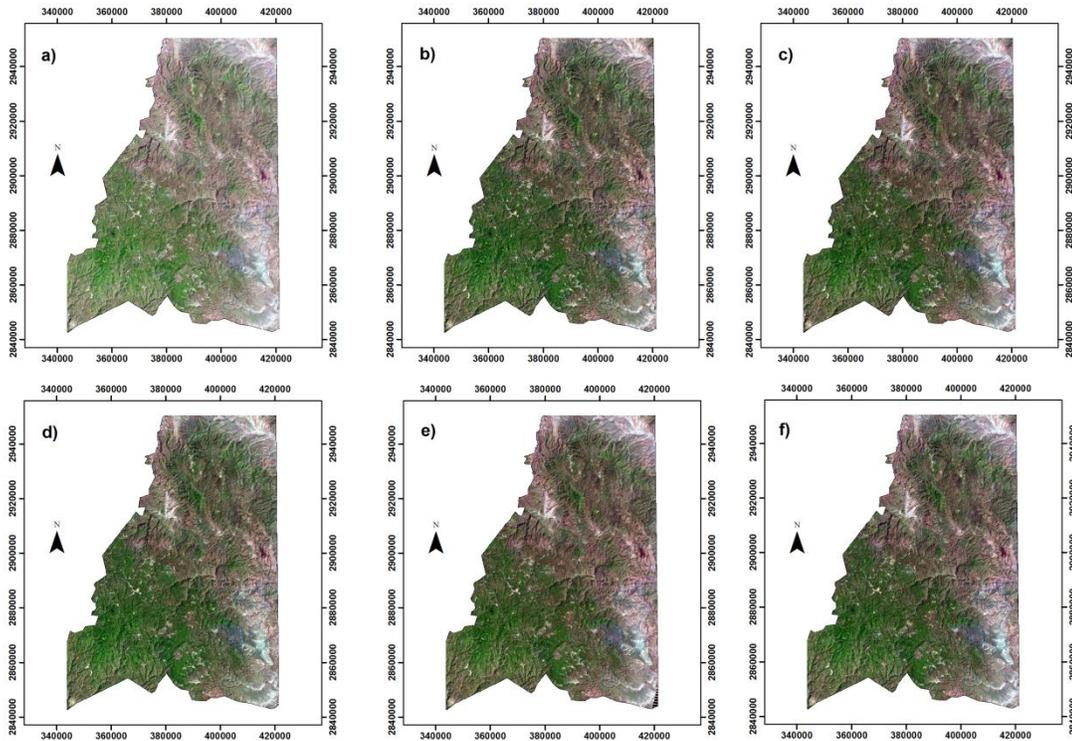
Figure 3. Comparison of the application of the different radiometric correction algorithms: (a) original image in DNs (multiband composition 5, 4 and 3), (b) ATCOR2, (c) COST, (d) FLAASH, (e) 6S, and (f) ToA.

The mean reflectances of the 99 SPIFyS plots obtained for each band of the Landsat 5 TM sensor and radiometric correction algorithm were graphically compared (Figure 4). The differences vary according to the atmospheric effects corrected by each particular algorithm. In general, the spectral signature of the vegetation showed in all cases typical behavior: low reflectance in the bands of the visible spectrum (450–690 nm) and high reflectance in the near- and mid-infrared bands (760–2350 nm). However, a clear difference appears depending on whether the radiometric correction algorithms were applied to variables with physical meaning (e.g., ground reflectance) rather than to DNs. In this respect, the ATCOR2 and FLAASH algorithms overcorrected the atmospheric effects (underestimation of transmissivity) in the short wavelength regions and underestimated them in the near- and mid-infrared regions, while COST and ToA produced the inverse spectral pattern. Finally, the performance of 6S was intermediate between that of the other algorithms evaluated.
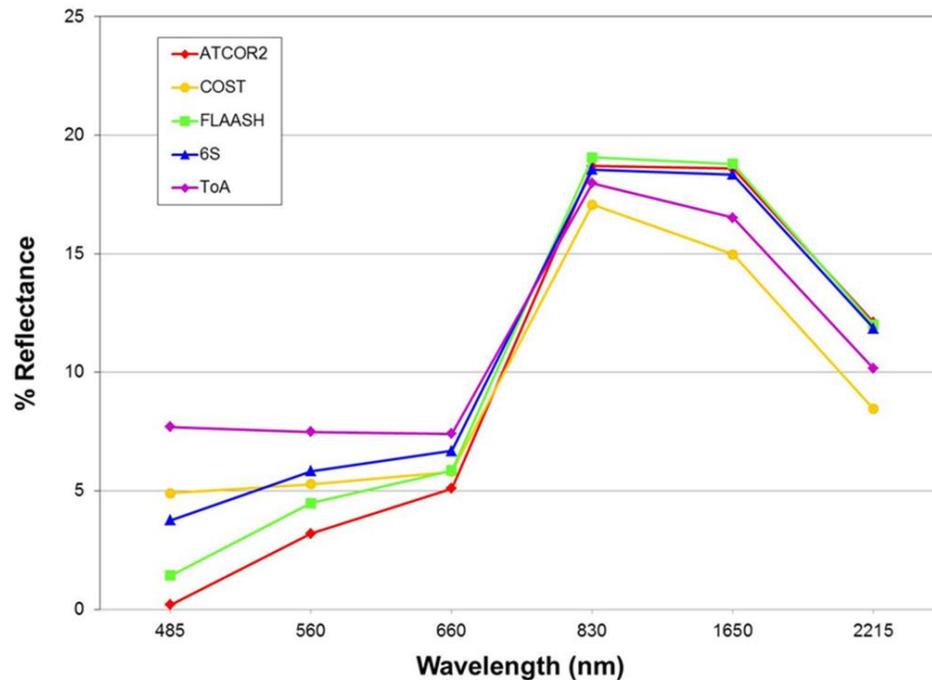
Figure 4. Spectral performance of each of the radiometric correction algorithms considered.

As the Shapiro–Wilk's test indicated that the corrected values of the spectral bands of the sensor were not normally distributed, therefore a log transformation was used. The ANOVA applied to the data from the 99 SPIFyS plots detected significant differences for all algorithms in the visible region (Figure 5), except for the COST and FLAASH algorithms in *Band 3*. Significant differences were observed in the near-infrared region for all algorithms tested, except ATCOR2 and 6S, which were included in the same population group (Tukey's test), they also shared characteristics with FLAASH and ToA. Finally, in the bands corresponding to mid-infrared regions, the algorithms ATCOR2, FLAASH and 6S did not provide significantly different results ($p > 0.05$).
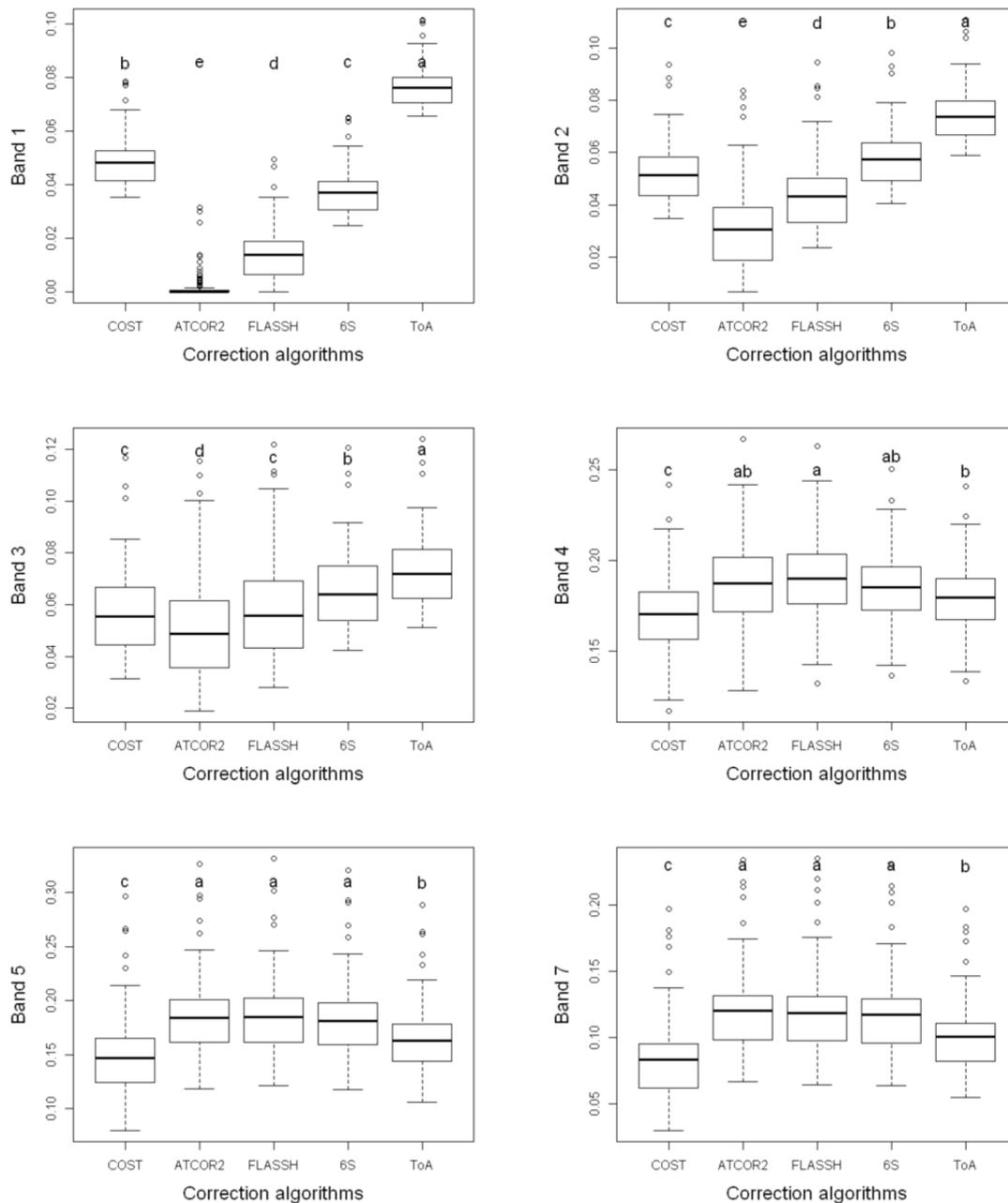
Figure 5. Box plots for each band of the Landsat 5 TM sensor and corresponding groupings of the radiometric correction algorithms (different letters indicate significant differences between algorithm performances, at $p < 0.05$).

The importance of the predictor variables (spectral bands and indices and topographic variables) for each of the MARS models fitted to the different radiometric correction algorithms was calculated (Figure 6). Results pointed out the near-infrared as the most relevant spectral region for AGB prediction. In fact, among the predictor variables, *Band*

7 made the greatest contribution to the capacity for predicting AGB for the different algorithms analyzed in almost all the cases except in algorithm 6S, where the variable *Band 5* appeared as the most important in accordance with the number of times that each variable is included in a subset.
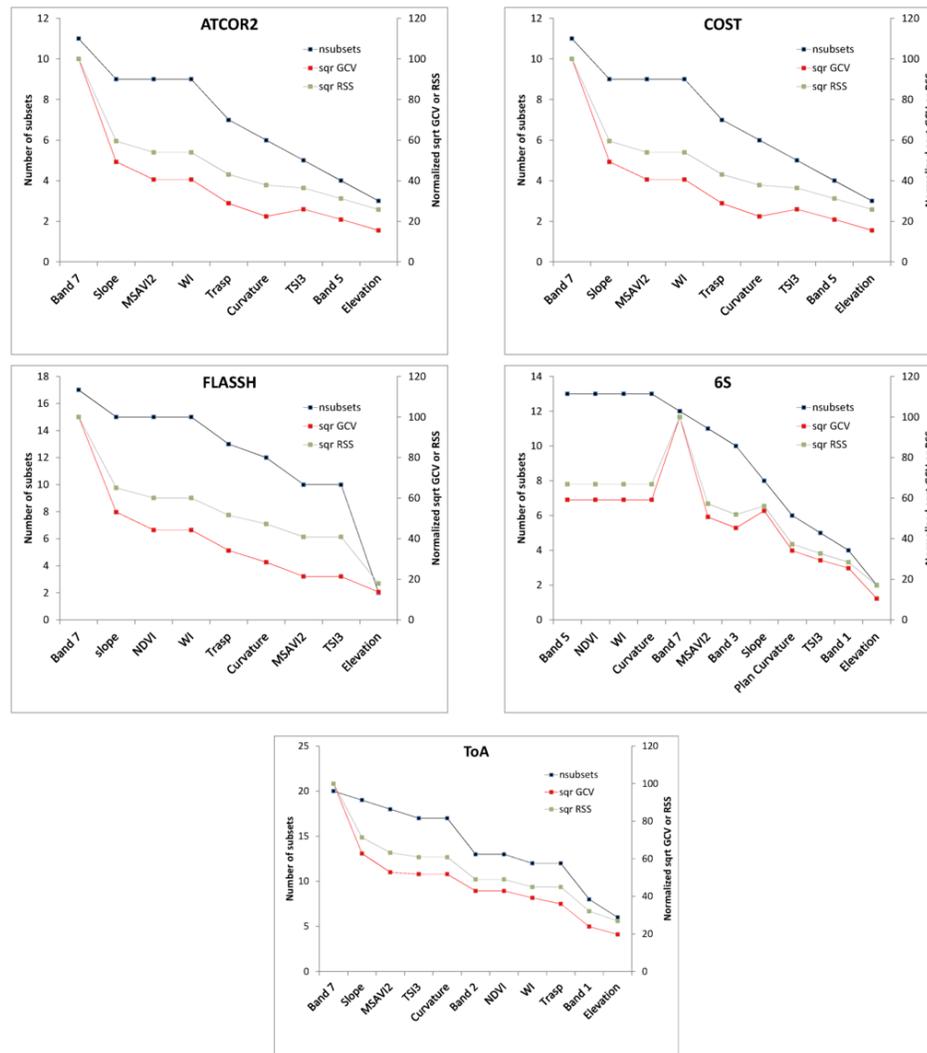
Figure 6. Importance and selection of predictor variables in the multivariate adaptive regression splines (MARS) models for each radiometric correction algorithm considered.

Table 3 shows the results of applying the MARS technique for estimating the AGB on the basis of different correction algorithms. The algorithm that yielded the best predictions was ToA ($R^2 = 0.89$, $RMSE = 29.82$ Mg·ha^{-1}, 33.49% of the average stand biomass).

However, after the application of cross-validation, the generalized coefficient of determination (GCV) decreased to 0.68. By contrast, the algorithm with the lowest predictive power was ATCOR2 ($R^2 = 0.75$ and $GR^2 = 0.59$). The algorithm with the highest generalized coefficient of determination was 6S ($GR^2 = 0.71$).

Table 3. MARS model selection criteria for AGB estimation and the different radiometric correction algorithms considered.

Algorithm	Number of terms	Number of predictors	GCV	RSS	R^2	GR^2	RMSE	%RMSE
ATCOR2	12 of 31	9 of 17	780.84	45556.61	0.75	0.59	50.37	56.58
COST	16 of 29	9 of 17	572.13	26722.49	0.85	0.70	36.55	41.05
FLAASH	18 of 33	9 of 17	737.91	30530.20	0.83	0.61	36.43	40.92
6S	16 of 30	12 of 17	552.26	25794.27	0.86	0.71	33.48	37.61
ToA	21 of 30	11 of 17	601.91	20452.83	0.89	0.68	29.82	33.49

where:
GCV: Generalized Cross Validation
RSS: Residual Sum of Squares
R^2: Coefficient of determination
GR^2: Coefficient of determination of GCV
RMSE: Root Middle of Squared of Error

Figure 7 shows the variations in the goodness of fit statistics (R^2 and GR^2) obtained by application of the cross-validation technique ($nfold = 2\text{–}10$) to the MARS model fits to each of the algorithms considered. In this case, the COST algorithm produced the best results and the highest stability, followed by ToA and 6S.
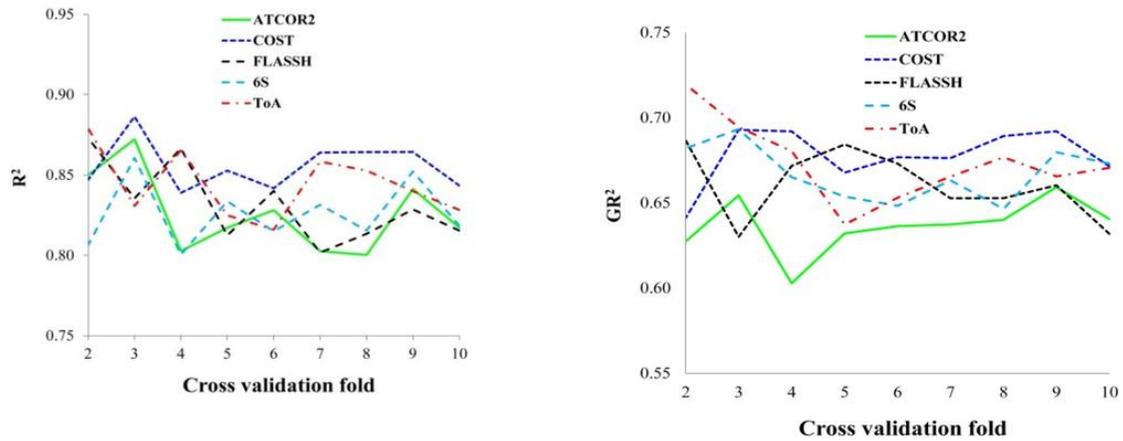
Figure 7. Variations in the goodness of fit statistics obtained by applying the cross-validation technique (*nfolds* = 2–10) to the MARS models.

Finally, AGB maps of the study area were generated by implementing the fitted MARS models for each radiometric correction algorithm on the satellite image using the raster calculator with ArcGIS® software (ESRI, 2011) and then intersected with the vector layer of vegetation cover (INEGI, 2012) (Figure 8).
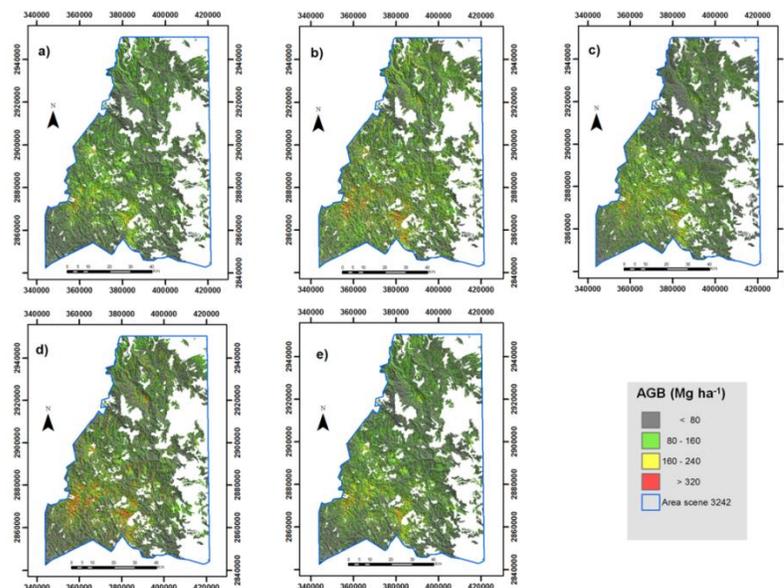
Figure 8. Maps of AGB in the study area generated from images corrected using the different radiometric correction algorithms considered. (**a**) ATCOR2, (**b**) COST, (**c**) FLAASH, (**d**) 6S, and (**e**) ToA.

Discussion

The use of radiometric correction algorithms enabled the transformation of DNs to reflectance values (*i.e.*, variables with biophysical meaning). The spectral reflectance of the forest cover in the study area was low (<0.10) in the visible region being typical of vegetated land (as chlorophyll absorbs most of the light received from the visible spectrum). Accordingly, the reflectance was higher (around 19%) in the near-infrared region, indicating the contrast between these regions of the electromagnetic spectrum (Fernández and Rodenas, 1999; Chuvieco, 2010) typical of green vegetation.

In accordance with (Chuvieco, 2010), reflectance in the blue band (450–520 nm) was lower relative to the DNs and was distorted by the ToA algorithm, which assumes a transparent atmosphere over flat land and perfectly Lambertian surfaces. However, with the ATCOR2 algorithm, overcorrection at the shortest wavelengths in the visible region (*i.e.*, green or red bands) is a consequence of scattering caused by particles of ozone and water vapor present in the atmosphere (Buho *et al*., 2009; Richter and Schläpfer, 2015). According to Broszeit and Ashraf (Broszeit and Ashraf, 2013), the images corrected using the COST algorithm are less accurate than those corrected with the ATCOR2 algorithm for vegetation cover, because COST automatically selects the lowest pixel values for each band to eliminate atmospheric haze from the data, whereas ATCOR2 uses specific parameters of atmospheric geometry and of the sensor for improved conversion of the reflectance data.

In the present study, the 6S algorithm performed more consistently than the other algorithms evaluated (ATCOR2, COST, FLAASH and ToA), because of the capacity of 6S to correct the effects of water vapor, high temperature and, therefore, high aerosol levels. This finding is similar to that reported by Nazeer *et al*. (Nazeer *et al.,* 2014), who observed that, among five algorithms tested, 6S produced the smallest difference in field-measured surface reflectance and that obtained using the Landsat ETM sensor.

The analysis of variance applied to data from the 99 SPIFyS plots revealed similar behavior of the population means for the ATCOR2, FLAASH and 6S algorithms in the bands of the mid-infrared region. However, the bands in the visible and near-infrared regions indicated a statistically-significant difference in the configuration of the groups of algorithms. These findings are consistent with those reported by Mahiny and Turner (Mahiny and Tuner, 2007), who compared the first four bands of the Landsat TM sensor

under five radiometric correction methods (two relative approaches: pseudo invariant features (PIF) and radiometric control sets (RCS), and three absolute approaches: COST, 6S and ToA), showing that in most cases, the four bands produced significantly different results.

Regarding the use of the MARS technique for estimating AGB in the present study, the ToA algorithm initially showed the greatest predictive power (R^2 = 0.89, RMSE = 29.8 Mg·ha^{-1}, 33.49% of the mean biomass in the stand). This result showed a slightly better performance than that reported by Hamdan *et al.* (2014) who estimated AGB by fitting regression models to reflectance data (corrected by ToA) from the SPOT-5 and ALOS PALSAR (R^2 = 0.803, RMSE = 32.6 Mg·ha^{-1}). Furthermore, in a study in southwest Colorado (USA), a lower correlation for biomass prediction (r = 0.86, RMSE = 45.6 Mg·ha^{-1}) with data corresponding to vegetation indices and texture analysis in Landsat TM images corrected by ToA radiometric correction was obtained (Kelsey and Neft, 2014). However, further analyses in the present study, such as the use of the generalized error of cross-validation (*GCV*), indicated a GR^2 value of 0.68, pointing out a decrease in the predictive capacity of the model. Therefore, after application of *GCV*, the 6S algorithm proved to be optimal (GR^2 = 0.71) despite the potentially negative effect of considering a larger number of predictors (12 of 17) than in ToA correction. It should be taken into account that the *GCV* criterion considers both the residual error and also the model complexity, penalizing those models with a high number of parameters. The results are similar to those obtained by Nguyen *et al.* (2015) in a study carried out in South Korea in which it was confirmed that the 6S algorithm fitted by the kNN technique (RMSE = 22.5 Mg·ha^{-1}) produced better results for estimating AGB than the other algorithms tested (DOS, FLAASH and ToA), especially for Landsat ETM bands in the infrared region. Furthermore, various studies have concluded that the MARS technique is a flexible method that yields robust predictions (Bilgili et al., 2010; Ghasemi and Zolfonoun, 2013).

The scatterplot of observed *versus* predicted aboveground forest biomass obtained with the MARS model fitted to the spectral data corrected using the proposed algorithm (6S) is shown in Figure 9. No trends to under- or over-estimate were observed.
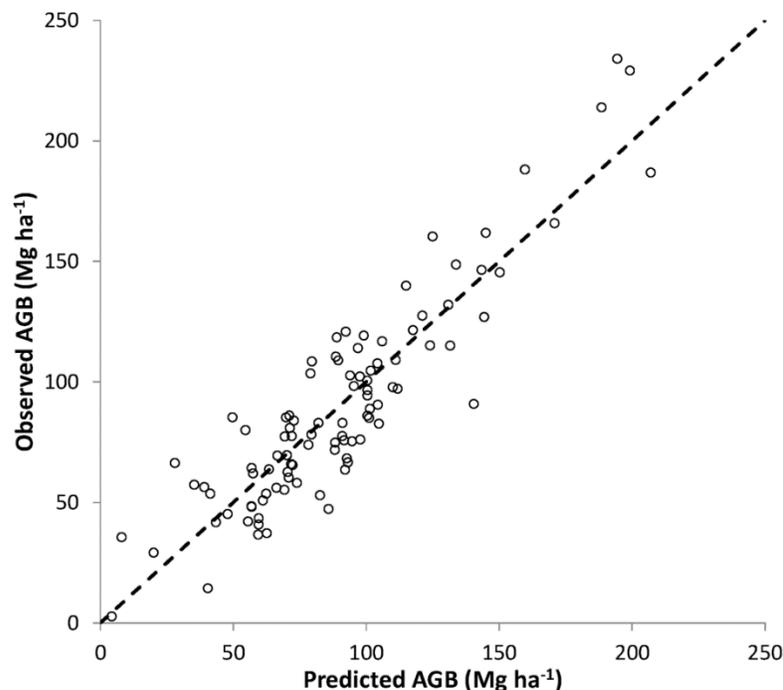
Figure 9. Observed *versus* predicted AGB obtained using the MARS model fitted to the spectral data corrected using the 6S algorithm. The broken line corresponds to the diagonal.

This statistical approach also enables identifying the importance of the predictive variables, in this case the near-infrared being the most important. Hence, in accordance with the *nsubsets* criterion, the spectral variable *Band 7* (2080–2350 nm) contributed most to the fitted MARS models for the ATCOR2, COST, FLAASH and ToA algorithms, in contrast to algorithm 6S, in which *Band 5* (1550–1750 nm) was the most important variable. However, the importance of the variables in algorithm 6S, under the criteria *sqr_rss* and *sqr_gcv*, indicates that *Band 7* makes the greatest contribution to the MARS model, as also occurred with the other algorithms tested. In this respect, the variable *Band 7* is associated with the moisture content of the forest stand, so that the greater the moisture content of the land cover, the greater absorption in this band of the electromagnetic spectrum, in other words, the surface reflectance captured by the sensor tends to decrease (Garcia *et al.*, 2005; López-Serrano *et al.*, 2015).

Among the topographical variables considered, the slope (β) and the wetness index (*WI*) contributed most to defining the MARS models and is also directly related to the moisture content in the field sites as a factor that controls run-off, the lower the value of β, the higher the moisture content and, therefore, the higher the AGB (Lam, 2004; Hoechstetter *et al.*,

2008; Kelsey and Neff, 2014). Hence, it is quite clear that radiometric and topographic variables related to water availability played a key role in the prediction of AGB in Durango's temperate forests.

Regarding the inclusion of vegetation indices, *NDVI* and *MSAVI2* were the variables that contributed most to the MARS models, given their potential usefulness in estimating AGB. Inclusion of these indices improved the prediction of the biophysical variables in forest ecosystems: in the case of *NDVI* because of it high sensitivity to the chlorophyll content of the vegetation (Sesnie *et al*., 2012; Ali *et al*., 2015) and in the case of *MSAVI2* because of its capacity to discriminate soil from vegetation, even in zones with scarce vegetation cover (Yao *et al*., 2015).

Conclusions

The study findings demonstrate the potential of the combined application of Landsat Thematic Mapper (TM) sensor satellite imagery and the consideration of topographical variables for estimating AGB, as well as the effects of different radiometric correction algorithms on the estimates obtained. The algorithm that showed the greatest spectral stability and that best estimated the AGB after application of the cross-validation technique was the 6S algorithm. However, each algorithm has advantages and disadvantages, and the user must parameterize each algorithm on the basis of the specific objectives. The MARS statistical method proved suitable for AGB estimation by using easy-to-obtain variables from remote sensing techniques. For all algorithms considered, the spectral band in the mid-infrared region (*Band 7*), the slope (*β*) and the wetness index (*WI*) of the land, both related to water availability, along with the vegetation indices *MSAVI2* and *NDVI* were the most important predictor variables for estimating AGB. Generation of an AGB map from the fitted models may be used in local and regional forest management, thus facilitating the location and precise delimitation of zones with different levels of AGB. Further research could aim at testing the effects of radiometric corrections on biomass modelling on a multi-scene and/or multi-temporal basis.

Forestal en el Norte de México) and by the research project EM2014-003 (Plan Galego de Investigación, Innovación e Crecemento 2011–2015, Consellería de Cultura, Educación e Ordenación Universitaria. Xunta de Galicia).

Author Contributions

Pablito M. López-Serrano and Carlos A. López-Sánchez conceived of, designed and performed the experiments and wrote the manuscript. Jose J. Corral-Rivas, Ramón A. Díaz-Varela and Juan G. Álvarez-González analyzed the data and revised the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

Adler-Golden, S., Berk, A., Bernstein, L.S., Richtsmeier, S., Acharya, P.K., Matthew, M.W., Anderson, G.P., Allred, C.L., Jeong, L.S., Chetwynd, J.H. 1998. FLAASH, a MODTRAN4 atmospheric correction package for hyperspectral data retrievals and simulations. In Procedings of the 7th Annual JPL Airborne Earth Science Workshop, Pasadena, CA, USA, 12–16 January 1998, Green, R.O., Ed., pp. 9–14.

Ali, M., Montzka, C., Stadler, A., Menz, G., Thonfeld, F., Vereecken, H. 2015. Estimation and validation of rapideye-based time-series of leaf area index for winter wheat in the rur catchment (Germany). Remote Sens. 7: 2808–2831.

Anderson, G.P., Felde, G.W., Hoke, M.L., Ratkowski, A.J., Cooley, T.W., Chetwynd, J.H., Jr., Gardner, J.A., Adler-Golden, S.M., Matthew, M.W., Berk, A., et al. 2002. MODTRAN4-based atmospheric correction algorithm: FLAASH (fast line-of-sight atmospheric analysis of spectral hypercubes). In Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII (Proceedings of SPIE), Shen, S.S., Lewis, P.E., Eds., Society of Photo Optics: Orlando, FL, USA, 2002, pp. 65–71.

Balthazar, V., Veerle, V., Eric, F.L. 2012. Evaluation and parameterization of ATCOR3 topographic correction method for forest cover mapping in mountain areas. Int. J. Appl. Earth Obs. Geoinform. 18: 436–450.

Baret, F., Guyot, G. 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sens. Environ. 35: 161–173.

Bilgili, A.V., Van Es, H.M., Akbas, F., Durak, A., Hively, W.D. 2010. Visible-near infrared, reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. J. Arid Environ. 74: 229–238.

Broszeit, A., Ashraf, S. 2013. Using different atmospheric correction methods to classify remotely sensed data to detect liquefaction of the February 2011 earthquake in Christchurch. In Proceedings of the SIRC NZ 2013—GIS and Remote Sensing Research Conference, Dunedin, New Zealand, 29–30 August 2013.

Buho, H., Kaneko, M., Ogawa, K. 2009. Correction of NDVI calculated from ASTER L1B and ASTER (AST07) data based on ground measurement. In Advances in Geoscience and Remote Sensing, Jedlovec, G., Ed., In-Tech Press: Vienna, Austria, 2009.

Carlotto, M.J. 1999. Reducing the effects of space-varying, wavelength-dependent scattering in multispectral imagery. Int. J. Remote Sens.20: 3333–3344.

Chavez, P.S. 1992. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sens. Environ. 1988, 24, 459–479.

Chavez, P.S. 1996. Image-based atmospheric corrections-revisited and improved. Photogramm. Eng. Remote Sens. 62: 1025–1036.

Chuvieco, E. 2010. Teledetección Ambiental: La Observación de la Tierra, 1st ed., Ariel Ciencia: Madrid, Spain.

Corral-Rivas, J.J., Vargas, B., Wehenkel, C., Aguirre, O., Álvarez, J.G., Rojo, A. 2009. Guía para el Establecimiento de Sitios de Inventario Periódico Forestal y de Suelos del Estado de Durango, Facultad de Ciencias Forestales: Universidad Juárez del Estado de Durango, Durango, México.

Cortés, L., Hernández, J., Valencia, D., Corvalán, P. 2014. Estimation of above-ground forest biomass using Landsat ETM+, Aster GDEM and Lidar. Forest Res.

Ediriweera, S., Pathirana, S., Danaher, T., Nichols, D., Moffiet, T. 2013. Evaluation of different topographic corrections for Landsat TM data by prediction of foliage projective cover (FPC) in topographically complex landscapes. Remote Sens. 5: 6767–6789.

ERDAS Inc. 2014. Erdas Imagine. Available online: http://www.hexagongeospatial.com/products/ERDASIMAGINE/details.aspx (accessed on 6 October 2014).

ESRI Inc. ArcGIS Desktop: Release 10, Environmental Systems Research Institute: Redlands, CA, USA, 2011.

EXELIS Inc. ENVI® 5.1. 2013. Fast Line-of-Sight Atmospheric Analysis of Hypercubes (FLAASH), EXELIS Inc.: Boulder, CO, USA.

Fernández, S.C., Rodenas, A.Q. 1999. Teledetección: Avances y Aplicaciones, Diputación de Albacete: Albacete, Spain.

Filippi, A.M., Güneralp, I., Randall, J. 2014. Hyperspectral remote sensing of above-ground biomass on a river meander bend using multivariate adaptive regressionsplines and stochastic gradient boosting. Remote Sens. Lett. 5: 432–441.

Friedman, J.H. 1991. Multivariate adaptive regression spline. Ann. Stat.19: 1–141.

Gagliasso, D., Hummel, S., Temesgen, H. 2014. A comparison of selected parametric and non-parametric imputation methods for estimating forest biomass and basal area. Open J. For. 4: 42–48.

Gao, B.C., Montes, M.J., Davis, C.O., Goetz, A.F. 2009. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. Remote Sens. Environ. 113: S17–S24.

García, M.A., Pérez, C.F., De la Riva, J., Fernández, J., Pascual, P.E., Herranz, A. 2005. Estimación de la biomasa residual forestal mediante técnicas de teledetección y SIG en masas puras de Pinus halepensis y P. sylvestris. In Proceedings of the IV Congreso Forestal Español, Zaragoza, Spain, 26–30 September 2005.

Geosystems. 2013. Haze Reduction, Atmospheric and Topographic Correction. User Manual ATCOR2 and ATCOR3, Geosystems GmbH: Geneva, Switzerland.

Ghasemi, J.B., Zolfonoun, E. 2013. Application of principal component analysis-multivariate adaptive regression splines for the simultaneous spectrofluorimetric determination of dialkyltins in micellar media. Spectrochim. Spectrochimi. Acta Part A Mol. Biomol. Spectrosc.115: 357–363.

Goslee, S. 2015. Radiometric and Topographic Correction of Satellite Imagery. Package Landsat. Version 1.0.8. Available online: http://cran.r-project.org/web/packages/landsat/landsat.pdf (accessed on 18 May 2015).

Güneralp, I., Filippi, A.M., Randall, J. 2014. Estimation of floodplain aboveground biomass using multispectralremote sensing and nonparametric modeling. Int. J. Appl. Earth Obs. Geoinform. 33: 119–126.

Guyot, G., Gu, X.F. 1994. Effect of radiometric corrections on NDVI-determined from SPOT-HRV and Landsat-TM data.Remote Sens. Environ. 49: 169–180.

Hadjimitsis, D.G., Papadavid, G., Agapiou, A., Themistocleous, K., Hadjimitsis, M.G., Retalis, A., Michaelides, S., Chrysoulakis, N., Toulios, L., Clayton, C.R.I. 2010. Atmospheric correction for satellite remotely sensed data intended for agricultural applications: Impact on vegetation indices. Nat. Hazards Earth Syst. Sci. 10: 89–95.

Hamdan, O., Hasmadi, I.M., Aziz, H.K. 2014. Combination of SPOT-5 and ALOS PALSAR images in estimating aboveground biomass of lowland Dipterocarp forest. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Tomst, Russia, 7–11 April 2014.

Hantson, S., Chuvieco, E. 2011. Evaluation of different topographic correction methods for Landsat imagery. Int. J. Appl. Earth Obs. Geoinform. 13: 691–700.

Hijmans, R.J. 2015. Raster: Geographic Data Analysis and Modeling. R Package Version 2.4-20. Available online: http://CRAN.R-project.org/package=raster (accessed on 6 October 2015).

Hoechstetter, S., Walz, U., Dang, L.H., Thinh, N.X. 2008. Effects of topography and surface roughness in analyses of landscape structure–A proposal to modify the existing set of landscape metrics. Landsc. Online, 3, 1–14.

Huang, C., Yang, L., Homer, C., Wylie, B., Vogelman, J., DeFelice, T. 2014. At Satellite Reflectance: A First Order Normalization of LANDSAT 7 ETM+ Images.

Available online: http://digitalcommons.unl.edu/usgspubs/109/(accessed on 6 October 2014).

Instituto Nacional de Estadística Geográfica e Informática (INEGI). 2012. Uso del Suelo y Vegetación Escala 1:250 000 Serie V, Información Vectorial, Instituto Nacional de Estadística Geográfica e Informática: Ciudad de Mexico, Mexico.

Instituto Nacional de Estadística Geográfica e Informática (INEGI). 2014. México Continuo de Elevaciones Mexicano 3.0 (CEM 3.0). Available online: http://www.inegi.org.mx/geo/contenidos/datosrelieve/continental/descarga.aspx(accessed on 6 May 2014).

Jaelani, L.M., Matsushita, B., Yang, W., Fukushima, T. 2015. An improved atmospheric correction algorithm for applying MERIS data to very turbid inland waters. Int. J. Appl. Earth Obs. Geoinform. 39: 128–141.

Jensen, J.R. 1996. A remote sensing perspective. In Introductory Digital Image Processing, Prentice Hall: Englewood Cliffs, NJ, USA.

Ju, J., Roy, D., Vermote, E., Masek, J., Kovalsky, V. 2012. Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods. Remote Sens. Environ. 122: 175–184.

Kelsey, K.C., Neff, J.C. 2014. Estimates of aboveground biomass from texture analysis of landsat imagery. Remote Sens. 6: 6407–6422.

Lam, C.S.Y. 2004. Comparison of Flow Routing Algorithms Used in Geographic Information Systems. Ph.D. Thesis, University of Southern California, Los Angeles, CA, USA.

Lavreau, J. 1991. De-hazing landsat thematic mapper images. Photogramm. Eng. Remote Ssens. 57: 1297–1302.

Lee, C.S., Yeom, J.M., Lee, H.L., Kim, J.J., Han, K.S. 2015. Sensitivity analysis of 6S-based look-up table for surface reflectance retrieval. Asia Pac. J. Atmos. Sci. 51: 91–101.

Levene, H. 1960. Robust tests for equality of variances. In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, Olkin, I., Ed., Stanford University Press: Stanford, CA, USA, 1960, pp. 278–292.

Li, F., Jupp, D.L., Thankappan, M., Lymburner, L., Mueller, N., Lewis, A., Held, A. 2012. A physics-based atmospheric and BRDF correction for Landsat data over mountainous terrain. Remote Sens. Environ. 124: 756–770.

Liang, S., Fang, H., Morisette, J.T., Chen, M., Shuey, C.J., Walthall, C.L., Daughtry, C.S.T. 2002. Atmospheric correction of Landsat ETM+ land surface imagery. II. Validation and applications. IEEE Trans. Geosci. Remote Sens. 40: 2736–2746.

López-Serrano, P.M., López-Sánchez, C.A., Díaz-Varela, R.A., Corral-Rivas, J.J., Solís-Moreno, R., Vargas-Larreta, B., Álvarez-González, J.G. 2015. Estimating biomass of mixed and uneven-aged forests using spectral data and a hybrid model combining regression trees and linear models. iForest.in press.

Mahiny, A.S., Turner, B.J. 2007. A comparison of four common atmospheric correction methods. Photogramm. Eng. Remote Sens. 73: 361–368.

McNab, W.H. 1989. Terrain shape index: quantifying effect of minor landforms on tree height. Forest Sci. 35: 91–104.

Milborrow, S. 2015. Earth: Multivariate Adaptive Regression Splines. R Package Version 4.2.0. Available online: http://CRAN.R-project.org/package=earth (accessed on 6 October 2015).

Moisen, G.G., Frescino, T.S. 2002. Comparing five modelling techniques for predict-ing forest characteristics. Ecol. Model. 157: 209–225.

Moore, I.D., Nieber, J.L. 1989. Landscape assessment of soil erosion and non-point source pollution. J. Minn. Acad. Sci. 55: 18–24.

Moran, M.S., Jackson, R.D., Slater, P.N., Teillet, P.M. Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. Remote Sens. Environ. 41: 169–184.

Moreira, E.P., Valeriano, M.M. 2014. Application and evaluation of topographic correction methods to improve land cover mapping using object-based classification. Int. J. Appl. Earth Obs. Geoinform. 32: 208–217.

Myers, R.H. 1990. Classical and Modern Regression with Applications, PWS-Kent Publishing Company: Belmont, CA, USA.

Nazeer, M., Nichol, J.E., Yung, Y.K. 2014. Evaluation of atmospheric correction models and Landsat surface reflectance product in an urban coastal environment. Int. J. Remote Sens. 35: 6271–6291.

Nguyen, H.C., Jung, J., Lee, J., Choi, S.U., Hong, S.Y., Heo, J. 2015. Optimal atmospheric correction for above-ground forest biomass estimation with the ETM+ remote sensor. Sensors. 15: 18865–18886.

R Core Team. 2014. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. Available online: http://www.R-project.org/ (accessed on 18 March 2014).

Richter, R. 1996. A spatially adaptive fast atmospheric correction algorithm. Int. J. Remote Sens. 17: 1201–1214.

Richter, R. 1996. Atmospheric correction of satellite data with haze removal including a haze/clear transition region.Comput. Geosci. 22: 675–681.

Richter, R., Schläpfer, D. 2015. Atmospheric/Topographic Correction for Satellite Imagery. ATCOR 2/3 User Guide, Version 9.0.0., ReSe Applications Schläpfer: Langeggweg, Switzerland.

Roberts, D.W., Cooper, S.V. 1989. Concepts and techniques of vegetation mapping. In Land Classifications Based on Vegetation: Applications for Resource Management, Ferguson, D., Morgan, P., Johnson, F.D., Eds., General technical report INT-US Department of Agriculture, Forest Service, Intermountain Research Station (USA): Ogden, UT, USA., pp. 90–96.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., Harlan, J.C. 1974. Monitoring the Vernal Advancements and Retrogradation (Greeen Wave Effect) of Natural Vegetation, NASA/GFSC: Greenbelt, MD, USA.

Roy, D., Wulder, M., Loveland, T., Woodcock, C., Allen, R., Anderson, M. 2014. LANDSAT-8: Science and product vision for terrestrial global change research. Remote Sens. Environ. 145: 154–172.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., et al. 2001. Benchmark map of forest carbon stocks in tropical regions across three continents. Proc. Natl. Acad. Sci. USA. 108: 9899–9904.

Sesnie, S.E., Dickson, B.D., Rosenstock, S.S., Rundall, J.M. 2012. A comparison of Landsat TM and MODIS vegetation indices for estimating forage phenology in desert bighorn sheep (Ovis canadensis nelsoni) habitat in the Sonoran Desert, USA. Int. J. Remote Sens. 33: 276–286.

Shapiro, S.S., Wilk, M.B. 1965. An analysis of variance test for normality (complete samples). Biometrika. 52: 591–611.

Song, C., Woodcock, C.E., Seto, K.C., Lenney, M.P., Macomber, S.A. 2001. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects?. Remote Sens. Environ. 75: 230–244.

Tan, K.C., Lim, H.S., Matjafri, M.Z., Adbullah, K. 2012. A comparison of radiometric correction techniques in the evaluation between LST and NDVI in Landsat imagery. Environ. Monit. Assess. 184: 3813–3829.

Tian, X., Li, Z., Su, Z., Chen, E., Van der Tol, C., Li, X., Guo, Y., Li, L., Ling, F. 2014. Estimating montane forest above-ground biomass in the upper reaches of the Heihe River Basin using Landsat-TM data. Int. J. Remote Sens. 35: 7339–7362.

Tyagi, P., Bhosle, U. 2011. Atmospheric correction of remotely sensed images in spatial and transform domain. Int. J. Image Process. 5: 564–579.

United States Geological Survey. 2011. Available online: http://glovis.usgs.gov (accessed on 11 January 2011).

Vanonckelen, S., Lhermitte, S., Van Rompaey, A. 2013. The effect of atmospheric and topographic correction methods on land cover classification accuracy. Int. J. Appl. Earth Obs. Geoinform. 24: 9–21.

Vargas-Larreta, B. 2013. Estimación del Potencial de los Bosques de Durango Para la Mitigación del Cambio Climático. Modelización de la Biomasa Forestal. Proyecto FOMIX-DGO-2011-C01-165681, Comisión Nacional de Ciencia y Tecnología (CONACYT): Durango, Mexico.

Vermote, E.F., Tanre, D., Deuze, J.L., Herman, M., Morcette, J.J. 1997. Second simulation of the satellite signal in the solar spectrum, 6S: An overview. IEEE Trans. Geosci. Remote. 35: 675–686.

Vincent, R.K. 1972. An ERTS multispectral scanner experiment for mapping iron compounds. In Proceedings of the 8th International Symposium on Remote Sensing of Environment, Ann Arbor, MI, USA, 2–6 October 1972, pp. 1239–1247.

Wilson, J.P., Gallant, J.C. 2000. Digital terrain analysis. In Terrain Analysis: Principles and Applications, Wilson, J.P., Gallant, J.C., Eds., John Wiley & Sons, Inc.: New York, NY, USA., pp. 1–28.

Yao, Z., Liu, J., Zhao, X., Long, D., Wang, L. 2015. Spatial dynamics of aboveground carbon stock in urban green space: A case study of Xi'an, China. J. Arid Land. 7: 350–360.

Zevenbergen, L.W., Thorne, C.R. 1987. Quantitative analysis of land surface topography. Earth Surf. Proc. Landf. 12: 47–56.

CAPITULO 6.

CONCLUSIONES GENERALES

En el presente estudio se ha demostrado el potencial de las imágenes satélite del sensor Landsat-5 TM, junto con la consideración de variables topográficas, variables de textura derivadas de índices de vegetación e índices de vegetación para cuantificar la AGB, constatándose, además, la influencia de diferentes algoritmos de corrección radiométrica de las imágenes en los resultados de dicha estimación.

1. Los métodos estadísticos no paramétricos implementados en la presente tesis permitieron cuantificar la biomasa forestal aérea con una precision aceptable, de acuerdo a la literatura citada, mediante el empleo de variables de fácil obtención a partir de técnicas de teledetección.

2. El enfoque que combina el método de regresión no paramétrica de árboles de regresión y clasificación (CART) y el análisis de regresión lineal múltiple (MLR) mejora el rendimiento del método de regresión lineal múltiple (MLR) por sí mismo.

3. Los resultados obtenidos en esta tesis muestran que el método M5P es potencialmente útil para la estimación de la biomasa forestal. La información contenida en los canales del infrarrojo cercano y medio del sensor Landsat-5 TM demostró ser la más eficiente para discriminar o predecir la AGB.

4. La realización de correcciones atmosféricas junto con la combinación de variables relacionadas con la estructura del bosque (SR con variables ASG) puede ayudar a resolver sensiblemente los problemas de saturación del NDVI para altos valores de biomasa.

5. Los algoritmos de aprendizaje automático kNN, RF y SVM son herramientas poderosas para la cuantificación de la biomasa aérea forestal con los conjuntos de datos de teledetección, convirtiéndose todos ellos en alternativas viables y precisas al método clásico de MLR paramétrica.

6. La elección del método (algoritmos de aprendizaje) dependerá en gran medida de la capacidad del usuario para llevar a cabo la parametrización del algoritmo, dado que estas técnicas, especialmente SVMs no son fáciles de aplicar y requieren un cierto grado de especialización. Nuestros hallazgos indican que la SVM es la mejor alternativa para los expertos, mientras que RF representa un equilibrio entre un modelo de precisión y una facilidad de uso para los no expertos.

7. El algoritmo de corrección atmosférica y radiometrica que mostró la mayor estabilidad espectral y que mejor estimó la AGB después de la aplicación de la técnica de validación cruzada fue el algoritmo 6S. Sin embargo, cada algoritmo utilizado en este estudio presenta ventajas y desventajas, y el usuario deberá parametrizar cada algoritmo sobre la base de sus objetivos específicos.

8. La banda espectral del infrarrojo medio (*Band7)*, la pendiente del terreno (*β)*, el índice de humedad y los índices de vegetación *MSAVI2* y *NDVI* fueron las variables predictoras con mayor importancia en la estimación de la AGB para todos los algoritmos considerados.

9. La generación de una cartografía de la AGB derivada de los modelos ajustados en este trabajo puede ser utilizada en los planes de manejo forestal a nivel local y regional, facilitando la localización y delimitación precisa de las zonas con diferentes rangos de AGB.

RECOMENDACIONES PARA FUTURAS INVESTIGACIONES

En los estudios de cuantificación de biomasa aérea forestal mediante sensores remotos en bosques templados de la Sierra Madre Occidental es recomendable considerar la geolocalización de las parcelas en campo con un GPS (Global Positioning System) de precisión submétrica, con el fin de poder disminuir el error de localización de dichas parcelas en las imaganes de satélite y mejorar la fiabilidad y robustez de los modelos ajustados.

Se sugiere ampliar la base de datos actual incluyendo un mayor número de parcelas con varios inventarios para obtener modelos más representativos y poder aplicarlos para estimar la AGB a una escala nacional empleando sensores remotos.

También se suguiere hacer nuevos estudios usando fuentes alternativas de datos espectrales como pueden ser imágenes de muy alta resolución espacial y espectral o el uso de drones y/o LIDAR (a nivel local para posteriormente ser implementado con otro sensor como Landsat para una escala global), especialmente para el caso de bosques con estructuras verticales y horizontales heterogéneas, con un gran número de especies y con amplios rangos de edades, como son los bosques de México, además de orientarse a probar los efectos de correcciones radiométricas en el modelado de la biomasa cuando se cuenta con información multi-escena y/o de base multi-temporal.