

ESTIMACIÓN DE LA DISTRIBUCIÓN ESPACIAL DE BIOMASA AÉREA Y ÁREA BASAL CON IMÁGENES DE SATELITE Y DATOS DE CAMPO EN UNA COMUNIDAD DEL CENTRO DE MÉXICO

Rita Libertad ADAME CAMPOS^a, Adrián GHILARDI^a, Jean- François MAS^a

^aCentro de Investigaciones en Geografía Ambiental,
Universidad Nacional Autónoma de México (UNAM)

Antigua Carretera a Pátzcuaro No. 8701, Col. Ex-Hacienda de San José de La Huerta, CP 58190 Morelia
Michoacán México

radame@pmip.unam.mx

RESUMEN

Presentamos un enfoque que integra imágenes de satélite, variables ambientales y datos de campo para modelar la distribución espacial de biomasa aérea. El estudio se realizó en una comunidad del centro de México, donde se realiza extracción intensiva de madera. Se midieron los diámetros del arbolado de 52 parcelas de 0.05 ha para estimar área basal y biomasa aérea en campo. Se generaron variables dependientes a partir de una imagen de satélite SPOT 5 y una RapidEye y se seleccionaron las mejores comparando dos técnicas estadísticas (stepwise-AIC y análisis de importancia relativa). Se utilizó la regresión lineal múltiple para generar los modelos y se validaron con validación cruzada. Los resultados indicaron que se pueden obtener modelos de distribución de área basal y de biomasa aérea confiables utilizando el análisis de importancia relativa como herramienta de selección de variables dependientes.

Palabras clave: SPOT 5, Rapideye, stepwise-AIC, importancia relativa, bosque templado.

ABSTRACT

The aim of this study is to show how an integrated use of satellite imagery, environmental variables and field data helps to model the spatial distribution of basal area and aboveground biomass. The study was conducted in a temperated forest in central Mexico, where intensive timber extraction is carried out. We measured basal area and estimated biomass data, in 52 0.05-ha forest plots. Dependent variables were generated from a SPOT 5 satellite image and one RapidEye and the best dependent variables were selected by comparing two statistical techniques. Multiple linear regression was used to interpolate, and the models were validated using cross-validation. Our results indicated that the spatial distribution of basal area and aboveground biomass may reliably be estimated using relative importance analysis.

Keywords: SPOT 5, Rapideye, stepwise-AIC, relative importance, temperate forest.

1 INTRODUCCIÓN

La biomasa aérea es una de las variables clave para inferir el estado de conservación de los bosques, lo que su vez tiene implicaciones a diferentes niveles (global, regional, local). Surge la importancia de contar con estimaciones de biomasa aérea confiables (i.e. validadas) y con rangos acotados de incertidumbre.

Entre los diversos enfoques que existen para estimar biomasa usando percepción remota, el modelo de regresión lineal (Sousa et. al., 2015; Wani et. al., 2015; Tian et al., 2012; Xie et. al., 2009) ha sido uno de las técnicas más utilizadas para modelar la correspondencia entre datos de campos y datos espectrales por su fácil y bajo costo de implementación (Zhu & Liu, 2015).

Sin embargo, en muchos casos se parte de un gran número y diversidad de variables biofísicas y espectrales que deben ser relacionadas con los datos de campo y elegir el menor número de variables capaces de explicar la mayor cantidad de variación puede resultar un reto.

El método stepwise y el criterio de información Akaike (AIC) han sido utilizados para selección de variables, análisis factorial, modelos de regresión y modelos de series temporales (Vaglio Laurin et al., 2014; Hero et al., 2013; Ene et al., 2012; Tian et al., 2012; Xie et al., 2009), por ser fácilmente manejados y ampliamente extendidos a modelos más generalizados, como los modelos lineales generalizados, modelos no lineales y una distribución no normal de los datos (Yamashita et al., 2007).

Por otro lado, el análisis de importancia relativa (RI) ha sido utilizado para medir la contribución de las variables predictoras en modelos de regresión múltiple (Johnson & Lebreton, 2004). Sin embargo, RI se puede utilizar para reducir los datos e identificar factores clave en la predicción y construcción de un modelo más parsimonioso (Lipovetsky & Conklin, 2015).

En este estudio se modeló la distribución de área basal y biomasa aérea de bosque templado en una comunidad del centro de México, comparando tanto el ajuste de los modelos como su desempeño (i.e. validación); y utilizando dos imágenes con diferente resolución espacial.

2 MATERIALES Y MÉTODOS

2.1 Área de Estudio

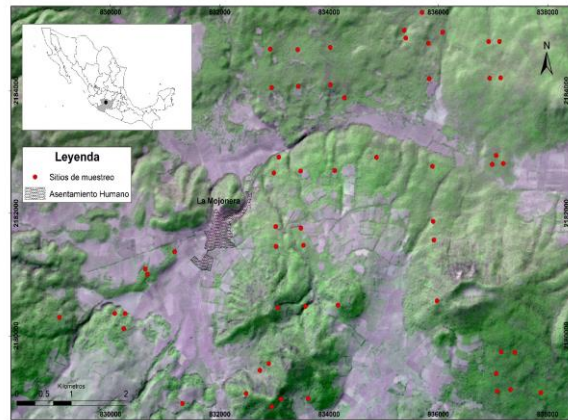
El estudio se realizó en una comunidad del estado de Michoacán, México entre $19^{\circ} 40' 30''$ y $19^{\circ} 44' 00''$ N, y $101^{\circ} 46' 30''$ y $101^{\circ} 52' 00''$ O, a una altura de 2300 y 3300 msnm sobre el Eje Neovolcánico Transversal.

2.2 Datos de Campo

Se obtuvieron datos de campo en 52 sitios georreferenciados de tipo circular anidado de 500

m² (radio de 12.62 m) **Error! Reference source not found.** Se recopilaron 13 variables para el estrato arbóreo (Morfín et. al. 2012) que caracterizan la estructura del bosque, que para este estudio dos variables resultaron relevantes: Área Basal (AB) y Diámetro a la Altura del Pecho (DN), que corresponde a 1.30 m. Se midieron los árboles > 5 cm DN y para la regeneración entre >2.5 cm y ≤5 cm DN.

Fig. 1 Localización de la comunidad de La Mojonera y sitios de muestreo



2.3 Variables independientes

Se utilizó una escena del sensor SPOT-5 HRG 2 pancromática y multiespectral, con fecha de 9 de febrero de 2014. Se ortorectificó previamente usando puntos de control utilizando como referencia ortofotos de la zona de estudio con una resolución espacial de 2 metros y el modelo de elevación de 15 metros de INEGI para eliminar errores de proyección. La ortorectificación se realizó con la herramienta Autosync Workstation de ERDAS 2011.

La escena del sensor Rapideye con una resolución espacial de 5 metros compuesta por 5 bandas (Azul: 440–510 m, Verde: 520 – 590 m, Rojo: 630 – 685 m, Rojo límite (EDG): 690 –730 m, Infrarrojo cercano (RNIR): 760 – 850 m), se utilizó de la fecha de 15 de febrero de 2013.

Las imágenes se corrigieron atmosférica y topográficamente con el módulo ATCOR3 de

ERDAS IMAGINE (Versión 11.0), un método de corrección radiométrica para terreno montañoso.

Para cada sitio se seleccionaron las variables independientes que explican la variación estructural del bosque a partir de la información recopilada en campo (Castillo-Santiago et. al. 2010). Para cada una de las imágenes, se exploraron las bandas espectrales, componentes principales, se generaron varios índices de vegetación Tabla 2, se calcularon 8 texturas GLCM (grey-level co-occurrence matrices; Mean, Std.Dev., Correlation, Dissimilarity, Entropy, ASM, Contrast and Homogeneity) en ventanas de 7x7 (SPOT 5) y 3x3(Rapideye) píxeles a partir de los valores de reflectancia (Halarick et. al. 1973, Zvoleff, 2015) y se derivaron variables del modelo de elevación (altitud, pendiente, orientación, radiación solar) siendo un total 56 variables derivadas de la imagen de satélite SPOT 5 y 68 para la imagen Rapideye.

2.4 Métodos estadísticos

Se analizó la relación entre las variables dependientes área basal (AB), biomasa (BIO), y las independientes con el coeficiente de correlación de Pearson para cada par (P-value de 0.05 y 0.01), para los índices de vegetación y texturas, se realizó el análisis por separado y se seleccionaron los coeficientes más altos para reducir el set de variables (Tabla 3).

Para seleccionar las variables explicativas idóneas, se consideró el método stepwise basado en el criterio de información Akaike (AIC) y el análisis de importancia relativa (RI). Ambos enfoques se implementaron en el software estadístico R con los paquetes 'stats' y 'relaimp' (R Core Team, 2015; Zvoleff, 2015).

En este sentido, se seleccionó el mejor modelo con el valor más bajo de AIC en función de las variables AB y BIO para las imágenes SPOT 5 y Rapideye.

La medida utilizada de importancia relativa en este estudio fue la descomposición marginal proporcional (PMD) para seleccionar las variables

independientes del mejor modelo de regresión. Se examinó la contribución de cada variable independiente con los valores más altos de varianza explicada en el modelo, se analizó el comportamiento de éstas al estar correlacionadas con otras y se descartaron las variables con los valores más bajos. Este procedimiento se realizó de forma manual hasta obtener el R^2 más alto y el menor error en el modelo de regresión.

Se aplicaron pruebas estadísticas de normalidad (Shapiro-Wilk) y homocedasticidad de los residuales (Breush-Pagan), se realizó el cálculo de los factores de inflación de varianza para detectar posibles problemas de multicolinealidad entre las variables independientes. A fin de eliminar heterocedasticidad o la falta de normalidad en los residuos, se analizaron transformaciones de las variables dependientes.

Se realizaron diagnósticos gráficos y distancia Cook para identificar observaciones con residuales altos o posibles apalancamientos y un análisis de regresión robusta (Li, 1985, Rousseuw, 1987).

Se utilizó la regresión lineal múltiple para interpolar los modelos finales y se validaron con la técnica LOOCV (leave-one-out cross-validation) evaluando su desempeño con el error cuadrático medio (RMSE) y el RMSE relativo (rRMSE) de la validación cruzada y considerando a los mejores modelos con el menor rRMSE (Efron, 1986).

3 RESULTADOS

3.1 Datos de Campo

En los 52 sitios se encontraron 2399 individuos, incluidos en 8 géneros (5 dominantes). El género que presentó valores más altos de área basal fue *Pinus sp* 44%, *Quercus sp.* 35%, *Acacia sp.* 15%, *Arbutus sp.* 3%, *Cedrus sp.* 2% y otros 1%. El 79% del área basal y biomasa aérea están representados por los 2 primeros géneros.

La tabla 2 muestra las estadísticas básicas de éstos parámetros.

	Promedio	Desviación Estándar	Varianza	Error Estándar	Mínimo	Máximo
Biomasa (Ton/ha)	116.94	66.38	4406.29	9.21	24.56	323.34
Area Basal (m ² /ha)	19.94	9.51	90.49	1.32	4.41	42.65

Tabla 1 Estadística Descriptiva de las variables muestreadas en campo.

3.2 Análisis de Correlación

La imagen de satélite Rapideye presentó los coeficientes de correlación más altos de AB y BIO.

Para la imagen SPOT 5 sólo la banda espectral infrarrojo medio (MIR) resultó significativa con correlación negativa, a diferencia de la imagen Rapideye en la que se obtuvieron coeficientes altos en las cuatro primeras bandas espectrales (Azul, Verde, Rojo y Borde Rojo) mostrando correlaciones negativas para las dos variables dependientes AB y BIO. Algunos componentes principales resultaron significativos para ambas imágenes, el NDVI resultó significativo para la variables AB en ambas imágenes de satélite, TVI para BIO de la imagen de satélite SPOT 5 (SPOT-BIO) y VI para BIO de la Rapideye (RAPIDEYE-BIO).

Las texturas que resultaron significativas fueron Media NIR para AB con la imagen de satélite SPOT 5 (SPOT-AB) y Second Moment NIR para SPOT-BIO así como Varianza RED para RAPIDEYE-BIO y Second Moment RED para AB de la imagen Rapideye (RAPIDEYE-AB). La elevación resultó significativa en ambas imágenes sólo para la variable BIO.

	SPOT-AB	SPOT-BIO	RAPIDEYE-AB	RAPIDEYE-BIO
Banda Azul	---	---	-0.471**	-0.489**
Banda Verde	---	---	-0.377**	-0.393**
Banda Roja	---	---	-0.508**	-0.511**
Borde Rojo	---	---	-0.279*	-0.323*
Banda Infrarrojo Medio (MIR)	-0.260*	---	---	---
Componente principal 1 (CP1)	---	---	-0.311*	-0.354**
Componente principal 2 (CP2)	-0.425**	-0.320*	0.284*	---
Componente principal 3 (CP3)	0.275*	---	-0.356**	-0.423**
Componente principal 4 (CP4)	-0.255*	---	---	---
Componente principal 5 (CP5)	---	---	0.451**	0.423**
NDVI	0.356**	---	0.545**	---
VI	---	---	---	0.535**
TVI	---	0.276*	---	---
Media NIR (7x7)	0.263*	---	---	---
Varianza RED (7x7)	---	---	---	-0.486**
Second Moment NIR (3x3)	---	-0.247*	---	---
Second Moment RED (3x3)	---	---	0.476**	---
Elevación	---	0.275*	---	0.275*

*P-value <=0.05
**P-value <=0.01

Tabla 2 Coeficientes de correlación para la relación lineal entre Área Basal y Biomasa comparadas con

variables (reflectancia espectral, texturas, índices de vegetación y topográficas) derivadas de las imágenes SPOT 5 y Rapid Eye.

3.3 Análisis de regresión

Para ambas imágenes, se identificaron falta de normalidad y homocedasticidad en el modelo de regresión con la variable respuesta original, mismas que fueron eliminadas transformando la variable a logaritmo natural. El análisis de valores atípicos y apalancamiento en las observaciones y el análisis de regresión robusta mostraron 3 observaciones que tenían un efecto notable en los ajustes de los modelos.

Las variables espectrales resultaron tener el mayor poder explicativo en ambas imágenes, particularmente la banda roja, borde rojo (sólo en Rapideye), algunos componentes principales y los índices de vegetación (NDVI, TVI y VI).

El modelo que presentó el coeficiente de determinación R² más alto (0.45) fue para la variable dependiente AB con variables dependientes derivadas de la imagen Rapideye y con el criterio de selección AIC (RAPIDEYE-AB-AIC).

Los modelos de regresión lineal seleccionados se muestran en las ecuaciones (1) y (2).

$$\ln(AB) = \beta_0 + \beta_1 \times NDVI \quad \text{Ec. 1}$$

$$\ln(BIO) = \beta_0 + \beta_1 \times CP5 + \beta_2 \times VI \quad \text{Ec. 2}$$

3.4 Validación de los modelos de regresión

Los resultados de la validación cruzada se muestran en la Tabla 5. Los modelos que presentaron el menor rRMSE para la variable dependiente AB fueron el generado por la imagen de satélite Rapideye con el análisis de importancia relativa (RAPIDEYE-AB-RI) con un rRMSE de 32.201 m²ha⁻¹ y para la variable BIO fue RAPIDEYE-BIO-RI con un rRMSE de 36.682 Ton ha⁻¹.

	MSE	RMSE	rRMSE (%)
SPOT-AB-AIC (m ² ha ⁻¹)	60.9	7.8	33.6
SPOT-AB-RI (m ² ha ⁻¹)	59.9	7.7	32.4
SPOT-BIO-AIC (Ton ha ⁻¹)	3223.6	56.8	38.8
SPOT-BIO-RI (Ton ha ⁻¹)	3223.6	56.8	38.8
RAPIDEYE-AB-AIC (m ² ha ⁻¹)	47.9	6.9	32.8
RAPIDEYE-AB-RI (m ² ha ⁻¹)	48.5	6.9	32.2
RAPIDEYE-BIO-AIC (Ton ha ⁻¹)	3277.8	57.3	38.7
RAPIDEYE-BIO-RI (Ton ha ⁻¹)	3147.7	56.1	36.7

Tabla 5. MSE, RMSE Y rRMSE de la validación cruzada de los modelos de regresión.

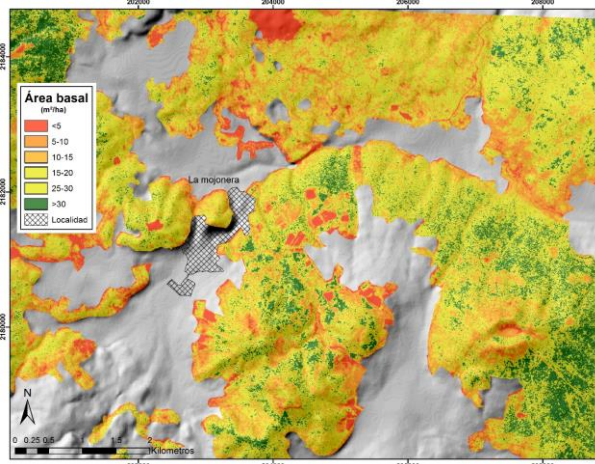


Fig. 3 Distribución espacial de área basal

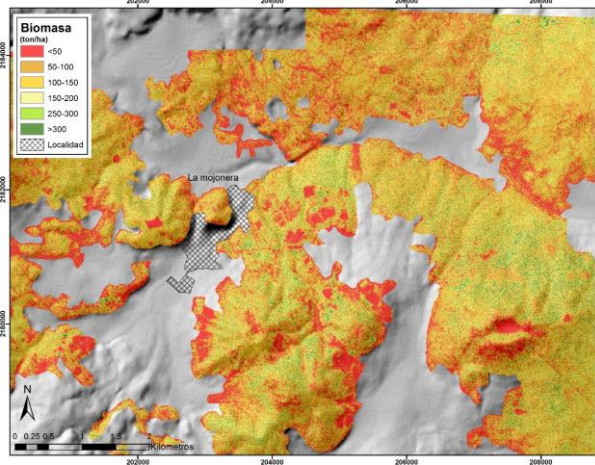


Fig. 4 Distribución espacial de biomasa aérea

3 DISCUSIÓN Y CONCLUSIÓN

La imagen de satélite que obtuvo el mejor ajuste fue la Rapideye para AB un R^2 de 0.40 y BIO un R^2 de 0.42. Los índices de vegetación (NDVI para AB y VI para BIO) fueron las variables

dependientes que presentaron los coeficientes de correlación más altos, ya que la respuesta espectral del NIR y Rojo está relacionada con la estructura celular de las plantas y las condiciones de humedad y para este estudio, los índices resultaron ser más sensibles a la cantidad de vegetación fotosintéticamente activa en el dosel (Kanemasu, 1974; Tucker, 1979).

Diversos estudios que predicen variables estructurales del bosque generalmente reportan el coeficiente de determinación R^2 y/o el RMSE como indicadores del ajuste de la relación entre las variables derivadas de imágenes de satélite y variables medidas en campo.

Muñoz-Ruiz (2014) estimó área basal con SPOT-5 y obtuvo un R^2 de 0.32 con un RMSE de 6.7 m² ha⁻¹ utilizando sólo variables espectrales e índices de vegetación. Wallner (2014), estimó área basal utilizando sólo variables espectrales y de textura derivadas de la imagen Rapideye y obtuvo un R^2 de 0.48 y un RMSE de 6.6 a 11.3 m² ha⁻¹, aplicando un enfoque de estratificación. Castillo-Santiago (2012), estimó biomasa con variables derivadas de imagen SPOT-5, obtuvo un R^2 de 0.69 y un RMSE de 9.77 t ha⁻¹, implementando análisis geoestadístico mejorando la capacidad predictiva del modelo. Kayitakire et al. (2006), obtuvo un R^2 de 0.35 y un RMSE de 6.8 m² ha⁻¹, con variables de textura derivadas de una imagen IKONOS para la variable dependiente área basal.

Aunque la mayoría de las variables dependientes potenciales no fueron significativas en relación con las variables independientes (AB y BIO); se obtuvieron ajustes de los modelos aceptables y con resultados similares a los resultados revisados, sin embargo podrían mejorar aumentando el tamaño de la muestra y aplicando un enfoque de estratificación por tipo de bosque (Lafiti et al. 2015, Wallner 2014).

Las diferencias entre los valores de rRMSE fueron mínimas; sin embargo, los modelos generados con el análisis de importancia relativa muestran para las dos imágenes de satélite errores menores y por lo tanto mejoran el desempeño predictivo de los modelos de regresión.

Se pueden obtener estimaciones aceptables para modelar la distribución espacial de parámetros estructurales forestales (área basal y biomasa aérea) utilizando en análisis de importancia relativa.

5 AGRADECIMIENTOS

Investigación realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT IA101513, UNAM. Análisis geoespacial de la degradación forestal por la extracción de madera para leña y carbón vegetal en el centro de México.

6 REFERENCIAS

- Efron, B. (1986). How Biased Is the Apparent Error Rate of a Prediction Rule ?, *81*(394), 461–470.
- Ene, L. T., Næsset, E., Gobakken, T., Gregoire, T. G., Ståhl, G., & Nelson, R. (2012). Assessing the accuracy of regional LiDAR-based biomass estimation using a simulation approach. *Remote Sensing of Environment*, *123*(2012), 579–592. doi:10.1016/j.rse.2012.04.017
- Hero, J. M., Castley, J. G., Butler, S. a., & Lollback, G. W. (2013). Biomass estimation within an Australian eucalypt forest: Meso-scale spatial arrangement and the influence of sampling intensity. *Forest Ecology and Management*, *310*, 547–554. doi:10.1016/j.foreco.2013.08.062
- Johnson, J. W., & Lebreton, J. M. (2004). History and Use of Relative Importance Indices in Organizational Research. *Organizational Research Methods*, *7*(3), 238–257. doi:10.1177/1094428104266510
- Lipovetsky, S., & Conklin, W. M. (2015). Predictor relative importance and matching regression parameters. *Journal of Applied Statistics*, *42*(5), 1017–1031. doi:10.1080/02664763.2014.994480
- R Core Team. (2015). R: A language and environment for statistical computing. The R Project for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Sousa, A. M. O., Cristina, A., Mesquita, P., & Marques, J. R. (2015). ISPRS Journal of Photogrammetry and Remote Sensing Biomass estimation with high resolution satellite images : A case study of *Quercus rotundifolia*, *101*(2015), 69–79. doi:10.1016/j.isprsjprs.2014.12.004
- 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*, *14*(1), 160–168. doi:10.1016/j.jag.2011.09.010
- Vaglio Laurin, G., Chen, Q., Lindsell, J. a., Coomes, D. a., Frate, F. Del, Guerriero, L., ... Valentini, R. (2014). Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *89*, 49–58. doi:10.1016/j.isprsjprs.2014.01.001
- Wani, A. A., Joshi, P. K., & Singh, O. (2015). Estimating biomass and carbon mitigation of temperate coniferous forests using spectral modeling and field inventory data. *Ecological Informatics*, *25*(2015), 63–70. doi:10.1016/j.ecoinf.2014.12.003
- Xie, Y., Sha, Z., Yu, M., Bai, Y., & Zhang, L. (2009). A comparison of two models with Landsat data for estimating above ground grassland biomass in Inner Mongolia, China. *Ecological Modelling*, *220*(15), 1810–1818. doi:10.1016/j.ecolmodel.2009.04.025
- Yamashita, T., Yamashita, K., & Kamimura, R. (2007). A Stepwise AIC Method for Variable Selection in Linear Regression. *Communications in Statistics - Theory and Methods*, *36*(13), 2395–2403. doi:10.1080/03610920701215639
- Zhu, X., & Liu, D. (2015). Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, *102*(2015), 222–231. doi:10.1016/j.isprsjprs.2014.08.014
- Zvloff, A. (2015). Calculate textures from grey-level co-occurrence matrices (GLCMs) in R. Retrieved from <https://cran.r-project.org/web/packages/g lcm/>