

EVALUACIÓN DE MÉTODOS DE ESTIMACIÓN PARA AJUSTAR ECUACIONES DE VOLUMEN EN PLANTACIONES DE TECA (*Tectona grandis* L.).

Argenis Mora Garcés y Ana Y. Moret

Universidad de Los Andes, Facultad de Ciencias Forestales y Ambientales, Instituto para el Desarrollo Forestal, Grupo de investigación Genética y Silvicultura, Mérida-Venezuela. E-mail: amora@forest.ula.ve

RESUMEN

En este trabajo se compararon tres métodos de estimación para ajustar el modelo alométrico simple $V_i = \beta_0(d_i^2 h_i)^{\beta_1}$ para predecir el volumen en plantaciones de teca (*Tectona grandis* L.), establecidas en la Unidad Experimental de la Reserva Forestal de Caparo, Barinas, Venezuela. Los tres métodos de estimación utilizados fueron el de distribución libre propuesto por Theil, un método de estimación No lineal y mínimos cuadrados ordinarios. Los métodos se evaluaron con datos provenientes de 174 árboles seleccionados mediante un muestreo estratificado por clases diamétricas. Los resultados sugieren que el modelo alométrico simple puede ser estimado utilizando el método de Theil y el clásico Mínimos cuadrados ordinarios con la transformación logarítmica en los datos. Sin embargo, en referencia a la validación cruzada, la estimación no lineal se comportó mejor para predecir volumen con corteza, mientras que para el volumen sin corteza, el método de Theil mostró mejor poder predictivo.

Palabras clave: *Tectona grandis* L., plantación, ecuaciones de volumen, regresión de Theil.

ABSTRACT

In this paper we compared three methods to fit the simple allometric model $V_i = \beta_0(d_i^2 h_i)^{\beta_1}$, in order to estimate individual volume in teak (*Tectona grandis* L.). A free distribution estimating method (Theil, 1950), non linear estimation and Least squares were used to fit data from 174 trees selected by stratified sampling considering diametrical classes as strata. The results suggest that the simple allometric model can be estimated by nonparametric method's Theil and Least square with log transformation in data; however, in reference to cross validation, nonlinear estimating had a good performance for volume with bark; therefore, the Theil method showed a very good performances for prediction of volumen with out bark.

Key words: *Tectona grandis* L, plantation, volume equations, Theil regression.

INTRODUCCIÓN

La estimación de la productividad y del rendimiento en plantaciones forestales es de gran importancia para su planificación y manejo sobre una base de sustentabilidad. En tal sentido, el volumen es considerado como una de las principales variables indicadora del potencial productivo del rodal. Las ecuaciones para estimar el volumen de árboles individuales suelen obtenerse a partir de modelos alométricos simples donde se relaciona el volumen de un árbol (V_i) con el diámetro medido a la altura de pecho (d_i) y altura total (h_i).

En el trabajo realizado por Moret (1997) se determinó, para plantaciones de teca (*Tectona grandis* L.) en los llanos occidentales de Venezuela, que el modelo alométrico simple:

$$V_i = \beta_0(d_i^2 h_i)^{\beta_1} \quad (1)$$

mostró el mejor ajuste para la estimación de volumen con y sin corteza, coincidiendo con Martínez *et al.* (1993) como ecuación general para diferentes especies en España. El modelo (1) representa un modelo intrínsecamente lineal y los residuos obtenidos a partir de él deben ser independientes y provenir de una distribución normal.

Las transformaciones tienen por objeto proporcionar un procedimiento más fácil de ajuste y/o procedimientos válidos de estimación y prueba. Por ejemplo, se puede convenir que la estimación del modelo anterior puede llevarse a cabo utilizando la transformación logarítmica; si esta se le aplica al modelo (1), se expresaría de la siguiente manera:

$$\text{Ln}(V_i) = \text{Ln}(\beta_0) + \beta_0 \text{Ln}(d_i^2 h_i) \quad (2)$$

Sin embargo, la estimación de V_i a partir de una función logarítmica utilizando directamente el anti-log del valor predicho de V_i tiende a subestimar, particularmente para valores de gran magnitud los cuales son comprimidos en la escala logarítmica (Beuchamp y Olson, 1973). Los autores precedentes basándose en las propiedades de la distribución lognormal, propusieron un método para obtener estimaciones insesgadas de variables transformadas con funciones logarítmicas, pero este requiere que la variable original (e.g. volumen por árbol) y los residuos sean aditivos y tengan una distribución normal; supuesto que no siempre pueda ser satisfecho, como por ejemplo, lo reporta Stromgaard (1986).

El presente trabajo compara y discute tres métodos de estimación, uno de distribución libre propuesto por Theil (1950), mínimos cuadrados ordinarios y un método de estimación no lineal aplicados sobre los modelos (1) y (2) para generar una ecuación de estimación para volumen con y sin corteza en una plantación de Teca.

MATERIALES Y MÉTODOS

Métodos de estimación

Theil (1950; citado por Hollander and Wolf, cap. 8, 1973) propuso un método de estimación no paramétrico del modelo de regresión simple; el procedimiento consiste en estimar β_1 de una línea de regresión por medio del cómputo de la mediana de las pendientes provenientes de todos los pares de conjuntos de puntos con diferentes valores de la variables regresora (X_i), es decir

$$\beta_{ij} = \frac{y_j - y_i}{X_j - X_i}, \text{ para } j > i, j = 1, 2, \dots, n$$

Se asume que todos los valores de X_i son diferentes y por conveniencia se ordenan los datos de manera ascendente en X , así existirá $n(n-1)/2$ distintos β_{ij} para un total de n pares de datos y el estimador Theil vendrá dado por:

$$\hat{\beta}_{\text{Theil}} = \text{mediana}(\beta_{ij})$$

Para el intercepto, Theil (1950) propone un estimador como la mediana de todos los interceptos posibles

$$\hat{\beta}_0 = y_i - \hat{\beta}_{\text{Theil}} X_i$$

La metodología de Theil (MT) sólo requiere que los residuos estén distribuidos independientemente con mediana 0 e independiente de X_i . Para la aplicación del método MT se definió a $\text{Ln}V_i$ como la variable y_i y como X_i a la variable combinada $d^2 h$.

Para la estimación de los parámetros del modelo (1) por regresión no lineal (NL) se utilizó el criterio de minimizar la diferencia entre la función estimada, llámese \hat{V}_i y la observada a través del criterio definido como función de pérdida (loss function)

$$\text{loss - function} = \sum_{i=1}^n (V_j - \hat{V}_i)^2 \quad (3)$$

cuanto menor sea esta función mejor será el ajuste. Para este caso se usó el algoritmo Gauss-Newton basado en las primeras y segundas derivadas de la función, usando las estimaciones mínimos cuadrados (MCO) del modelo transformado como valores de arranque (Ross, 1990, cap. 5). Para comparar la bondad de ajuste del modelo alométrico simple por los métodos de estimación, descritos anteriormente, se utilizaron los criterios: el coeficiente de determinación (R^2), el cuadrado medio de los residuos y las estimaciones puntuales y por intervalos de confianza al 95 % para el coeficiente de regresión β_1 . Nótese que este coeficiente puede ser estimado directamente ya que no es afectado por la transformación, mientras que para el intercepto es necesario aplicar anti-log.

Cada ecuación obtenida fue sujeta a una validación cruzada para conocer el potencial predictivo en ellos. Para ello, se utilizó una muestra independiente de 39 árboles y se compararon de acuerdo a los criterios: suma de cuadrados de los residuos de predicción (SCR pred), el coeficiente de determinación de predicción (R^2 pred) (Montgomery y Peck, 1981) y el error cuadrático medio (ECM) que combina la precisión y el sesgo de predicción en cada ecuación ajustada (Alexandros *et al.*, 1992), y se define como:

$$\text{ECM} = \bar{\epsilon}^2 + V \quad (4)$$

donde \bar{e} refiere al promedio de los residuos entre el volumen real y el predicho, V es un estimador de la varianza de predicción de los residuos en cada ecuación. Además de los criterios mencionados anteriormente, los valores \bar{e}^2 , \sqrt{V} y \sqrt{ECM} fueron usados también como criterios para comparar la validez en la predicción de las ecuaciones ajustadas por los tres métodos de estimación.

La muestra

El estudio se realizó con datos provenientes de plantaciones de *Tectona grandis* L, establecidas en 1973 a una altura de 100 msnm en la Unidad Experimental de la Reserva Forestal Caparo, localizadas en los Llanos Occidentales del Estado Barinas, Venezuela. Se aplicó un muestreo aleatorio estratificado considerando las clases diamétricas (con una amplitud de dos centímetros) presentes en la población como estratos y se recolectó información de diámetro a la altura de pecho (d, cm) y altura total (h, m). El volumen con y sin corteza fue estimado por la fórmula de Smalian (Alder, 1980) sobre un total de 174 árboles; para ello, se procedió a tumbar los árboles y se midió el diámetro cada 2 m hasta un diámetro mínimo aprovechable. Para obtener una muestra balanceada por categoría diamétrica, con un error permisible porcentual máximo del 10%, y previendo cubrir toda la variabilidad presente en cada una de esas categorías, se decidió cubicar 16 árboles por clase diamétrica.

RESULTADOS Y DISCUSIÓN

El cuadro 1 muestra las estadísticas descriptivas para las variables volumen con y sin corteza, diámetro (d) y altura total (h) utilizados para el ajuste de las ecuaciones, así como, para las variables volumen con y sin corteza para la validación.

Los resultados del ajuste del modelo alométrico (1) por mínimos cuadrados ordinarios (MCO), estimación no lineal (NL) y el método de Theil (MT) se muestran en el cuadro 2. Nótese como existe poca diferencia en las estimaciones del parámetro β_0 para el volumen con y sin corteza, respectivamente. Esto sugiere que la estimación de este parámetro no se ve afectada por del método de estimación utilizado. Al observar el comportamiento de los tres métodos respecto a la estimación del parámetro β_1 , se tiene que para la relación volumen con corteza - d^2h , la estimación por MCO rindió un intervalo de confianza (95 %) más preciso que sus competidores, a su vez, se obtuvo una subestimación de este coeficiente por estimación NL cuando se compara con los métodos MCO y MT; aunque con mayor precisión en la estimación general ($CMR = 22.242 \times 10^{-4}$).

Respecto a la relación volumen sin corteza, el método MT rindió una estimación por intervalos con una banda de confianza mucho menor; igualmente, la estimación NL comparativamente subestima el parámetro β_1 . Se debe notar que para la estimación no lineal, el mínimo en las sumas de cuadrados de los residuales se lleva a cabo directamente sobre V_i en el modelo (1), mientras que para el modelo (2), el mínimo se obtiene sobre $\ln V_i$. Al revisar la distribución de los residuales obtenidos por la estimación de cada método, se tiene que estos mostraron distribuciones sesgadas (prueba Shapiro-Wilk, $p < 0.01$) lo cual, en principio, se recomienda la ecuación ajustada por la metodología no paramétrica de Theil.

El cuadro 3 resume los resultados de la validación de las ecuaciones ajustadas por MCO, NL y MT para una muestra independiente de volumen real con y sin corteza, respectivamente, presentando los criterios \bar{e}^2 , \sqrt{V} y \sqrt{ECM} , sumas de cuadrados del error de predicción (SCE pred) y el coeficiente de determinación (R^2 pred). La ecuación estimada por el método MT generó predicciones ligeramente menos

Cuadro 1. Resumen de las estadísticas descriptivas para las variables diámetro (d), altura total (h), volumen con corteza (Vcc) y sin corteza (Vsc) utilizadas para evaluar los métodos de estimación y validación.

Estadísticas	d (cm)	h (m)	Vcc	Vsc	Vcc*	Vsc*
Min	6.5	8.8	0.021	0.014	0.061	0.053
Media	21.1	20.9	0.372	0.284	0.277	0.251
Max	27.7	35.8	1.184	0.916	0.594	0.549
CV %	20.1	31.6	66.2	67.3	51.6	53.4

*: Volumen con y sin corteza usados para la validación de las ecuaciones

Cuadro 2. Resultados de los ajustes del modelo alométrico por estimación mínimos cuadrados (MCO), No lineal (NL) y Theil (MT) para volumen con y sin corteza.

Volumen	Método de Estimación	b_0	li	b_1 puntual	ls	CME x10 ⁻⁴	R ²
Con Corteza	MCO	0.341127	0.918764	0.93590	0.953036	22.675	0.98521
	N L	0.348571	0.869404	0.902762	0.936120	22.242	0.96474
	MT	0.340291	0.915883	0.936207	0.967879	22.650	0.96409
Sin Corteza	MCO	0.259562	0.943381	0.961016	0.978651	13.573	0.98515
	N L	0.264909	0.884112	0.918298	0.952485	13.518	0.96444
	MT	0.258403	0.951654	0.961679	0.973714	13.992	0.96320

li y ls: límite inferior y superior respectivamente para el intervalo de confianza (95%), CME: Cuadrado medio del error, R²: Coef. Determinación.

Cuadro 3. Resultados de la validación independiente de las ecuaciones ajustadas por mínimos cuadrados, No lineal y Theil para volumen con y sin corteza.

Volumen	Método de Estimación	$\bar{\epsilon}$	\sqrt{V}	\sqrt{ECM}	SCE predx10 ⁻⁴	R ² pred
Con Corteza	Mínimos Cuadrados	- 0.0278	0.0521	0.0590	34.1271	0.829022
	No Lineal	- 0.0343	0.0517	0.0620	37.7750	0.811075
	Theil	- 0.0270	0.0519	0.0586	33.6343	0.831492
Sin Corteza	Mínimos Cuadrados	0.0189	0.0548	0.0586	33.6043	0.822104
	No Lineal	0.0143	0.0555	0.0574	32.1100	0.830014
	Theil	0.2003	0.0555	0.0590	34.0729	0.819623

sesgadas que sus contrapartes paramétricos MCO y NL; en promedio, se tiene que las tres ecuaciones produjeron subestimaciones respecto a los valores reales de volumen con corteza. La mejor combinación de mayor precisión y menor sesgo para la predicción del volumen con corteza se obtuvo para el método no paramétrico, reflejado también en los criterios R² pred y SCE pred.

La revisión del comportamiento de las ecuaciones ajustadas para predecir el volumen sin corteza indica, en promedio, que todas sobre estiman los valores reales. Sin embargo, la ecuación ajustada por estimación NL tuvo menor sesgo de predicción y a su vez, fue ligeramente superior en cuanto a la mejor combinación mayor precisión – menor sesgo ($\sqrt{ECM}=0.574$) y seguido de igual manera para los criterios R² pred y SCE pred. Nótese cómo la precisión en la predicción fue muy similar para las tres ecuaciones.

CONCLUSIONES

Se concluye que, para el conjunto de datos analizados en este estudio, el ajuste de la ecuación alométrica simple (1) puede ser estimada vía linealización por la transformación log (modelo transformado (2)) utilizando los métodos de estimación no para-métrico de Theil y el clásico Mínimos cuadrados ordinarios.

Como en este caso la distribución de los residuos no siguió una distribución normal, se recomienda la estimación noparamétrica propuesta por Theil.

Si bien la estimación no lineal tuvo un buen comportamiento predictivo, se prefieren a las estimaciones por mínimos cuadrados y el no paramétrico, por ser estos métodos de fácil aplicación.

En cuanto a la validación, el método Theil arrojó mejor comportamiento para la variable volumen con corteza, no obstante, para volumen sin corteza, el

método de estimación no lineal mostró mejor comportamiento desde el punto de vista predictivo.

STROMGAARD, P. 1986. Biomass estimation equations reviewed-the example from the Zambian Miombo. *Agroforestry System*, 4, 375-379.

AGRADECIMIENTO

Los autores agradecen al Consejo de Desarrollo Científico, Humanístico y Tecnológico (CDCHT) de la Universidad de Los Andes por el apoyo económico para el desarrollo de la presente investigación, la cual se desarrolló en el marco del proyecto Central del Grupo de Investigación Genética y Silvicultura (Código FO-337-95-01-A) del Instituto de Investigaciones para el Desarrollo Forestal, Facultad de Ciencias Forestales y Ambientales.

REFERENCIAS BIBLIOGRÁFICAS

- ABBOT, P.; J. LOWORE y M. WERREN. 1997. Models for estimation of single tree volume in four Miombo woodland types. *Forest ecology and Management*, 97, 25-37.
- ALDER, A. 1980. *Estimación del volumen forestal y predicción del rendimiento*. Volumen 2. F.A.O.
- ALEXANDROS, A. and H. BUNKHART. 1992. An evaluation of sampling methods and model forms for estimating height - diameter relationships in Loblolly Pine plantations. *Forest Science*, 38 (1): 192-198.
- BEUCHAMP, J. and J. OLSON. 1973. Correction for bias in regression estimates after logarithmic transformation. *Ecology*, 54: 1403-1407.
- MONTGOMERY D. and E. PECK. 1981. Introduction to linear regression analysis. John Willey and sons, New York. Pp. 446.
- MARTINEZ, J.; P. ARA y I. GONZALEZ. 1993. Ecuaciones alométricas de tres variables: estimación del volumen, crecimiento y porcentaje de corteza de las principales especies maderables españolas. *Investigación Agraria. Sistemas y recursos forestales*, 2: 211-228.
- MORET, A. 1997. *Ecuaciones de volumen en teca*. Facultad de Ciencias Forestales y Ambientales, Universidad de Los Andes. 61 pp.
- ROSS, G.J.S. 1990. *Non linear estimation*. Springer-Verlag New York, inc. 189p.
- HOLLANDER, M. and D. WOLF. 1973. *Non parametrical statistical methods*. New York. John Wiley and Sons. Pp. 200-218.