

## Modelos de volumen para *Pinus pinaster* Ait. en la Comarca del Baixo Miño, Galicia, España

*Volume equations for Pinus pinaster Ait. in the Baixo Miño region, Galicia, Spain*

Christian Salas<sup>1\*</sup>, Laura Nieto<sup>2</sup> y Alfonso Irisarri<sup>3</sup>

Recibido en septiembre de 2005; aceptado en diciembre de 2005

### RESUMEN

Se ajustaron ecuaciones que estiman el volumen de árboles individuales de *Pinus pinaster* Ait (Pino Gallego) en la comarca del Baixo Miño (Galicia, España). La obtención de la mejor ecuación de volumen se realizó en las siguientes cuatro etapas: prueba de cumplimiento de los supuestos estadísticos, ranking de los modelos de acuerdo a la bondad de ajuste, ranking de los modelos de acuerdo a sus capacidades predictivas, y análisis de las capacidades predictivas dentro de rangos de clases volumétricas. Los modelos ajustados fueron validados con una muestra independiente de árboles. El modelo seleccionado presenta una raíz del error cuadrático medio (RECM) de un 7,77% y un sesgo (DA) de 0,25%. Estos valores son sustancialmente menores que el error y sesgo obtenido al emplear la ecuación ajustada en los inventarios forestales nacionales para la provincia de Pontevedra, que emplea el mismo número de variables independientes (diámetro normal, altura total y el diámetro fustal a los 4 m de altura). El modelo propuesto, además fue el único que cumplió con los supuestos estadísticos que sustentan el análisis de regresión.

**Palabras clave:** tarifas de cubicación, análisis de regresión.

### ABSTRACT

Volume equations were fitted for individual trees of *Pinus pinaster* Ait (Maritime pine) at the Baixo Miño region in Galicia, Spain. To obtain the best volume equation the following four steps were taken: the test on statistical assumptions' satisfaction, model ranking according to their goodness of fitting, model ranking according to their prediction accuracy, and finally, an analysis of the models prediction capacity within the range of volumetric classes. The fitted models were validated using an independent sample dataset. The selected model showed a root of mean squared errors (RMSE) of 7.77% and a bias (DA) of 0.25%. These values are notably smaller than the error and bias obtained using the fitted model from the national forest inventories for the province of Pontevedra which uses the same number of independent variables (diameter at breast height, total height and upper-stem diameter at 4 m). In addition, the proposed model was the only one that satisfied the statistical assumptions that support the regression analysis.

**Key words:** volume equations, regression analysis.

<sup>1</sup> Departamento de Ciencias Forestales. Universidad de La Frontera. Casilla 54-D. Temuco. Chile.

<sup>2</sup> Montes y Medio Ambiente Consultores S. L. Rúa Michelena 1, 7º. 36.002 Pontevedra, España.

<sup>3</sup> FESERMAGA Rúa San Lázaro 93 – 2º A. 15703 Santiago de Compostela, España.

\* Autor para la correspondencia. Dirección actual: Department of Statistics, University of Idaho, PO Box 441104, Moscow ID 83843-1104, USA. Email: csalas@uidaho.edu. Tel. +1(208)885-5165.

Fax +1(208)885-6226. sala3886@uidaho.edu sala3886@uidaho.edu

## 1. INTRODUCCION

La Comarca del Baixo Miño comprende los términos municipales de Tui, Tomiño, A Garda, Oia y O Rosal, en Galicia, España. La superficie total que abarca es de 32.165 ha de las cuales 24.134 ha están consideradas como superficie forestal. Dentro de ésta la importancia de la especie *Pinus pinaster* es clara, ocupando el 43% de la superficie total de la comarca y un 56% del total forestal, así como también se proyecta una gran importancia en términos de producción de volumen (García, 2001).

Una de las actividades más importantes en biometría forestal es la estimación de volumen de árboles (Robinson y Wood, 1994). Las ecuaciones de volumen que predicen el volumen de árboles individuales en base a algunas variables dependientes de fácil medición en el campo (por ejemplo; diámetro a la altura del pecho “ $d$ ”, altura total “ $h$ ”), son herramientas fundamentales en nuestros días para la estimación de existencias en bosques, y a través de esto, se convierten en un punto básico en la toma de decisiones en gestión forestal, debido a que mayormente los bosques son valorados económicamente en función de la cantidad (volumen) de madera que estos producen.

La necesidad de disponer de ecuaciones que permitan la estimación de valores precisos de las existencias de una masa forestal es vital en manejo forestal. Si a esto se une el creciente interés por la planificación de la gestión de las masas forestales y por su ordenación, en definitiva, aumentan la importancia de contar con modelos de estimación de volumen (existencias).

La mayoría de los estudios sobre estimación de volumen para *Pinus pinaster* en Galicia, han sido desarrollado empleando una gran escala geográfica. Rodríguez (1995) ajustó modelos de volumen de árbol individual para la zona costera e interior de Galicia, proponiendo finalmente modelos dependientes del  $d$  y  $h$ . Diéguez *et al.* (2003) ajustaron modelos de volumen individual para *Pinus pinaster* empleando como variable predictora (independiente) el diámetro de tocón “ $d_{st}$ ”. Sin embargo, el uso de esta ecuación está centrado en la estimación del volumen de un árbol, cuando éste ha sido cortado. Recientemente, Rojo *et al.* (2003) ajustaron modelos de ahusamiento (funciones de perfil) para la especie en Galicia, pero su evaluación no contemplo la estimación de volumen, y el modelo finalmente propuesto por ellos requiere una alta manipulación computacional debido a procesos de iteración para la integración del modelo de ahusamiento, con la cual se puede estimar volumen, lo cual si bien es posible de realizar, en el uso de los modelos por usuarios forestales, resulta complejo.

Debido a lo anterior y al marcado carácter especie - región que presentan las masas forestales se ha elegido la comarca natural del Baixo Miño para focalizar un estudio sobre estimación de volumen individual en *Pinus pinaster*. El objetivo de este artículo es ajustar ecuaciones de volumen individual para *Pinus pinaster* en la comarca del Baixo Miño en Galicia, empleando un riguroso procedimiento estadístico de selección de modelos.

## 2. MATERIAL Y METODOS

### 2.1 Datos

El área donde se llevó a cabo la investigación abarca los términos municipales de Tui, Tomiño, A Garda, Oia y O Rosal, en Galicia, España. Los rodales donde se midieron árboles muestra, poseían en promedio 37 años de edad, 1070 arb/ha y 34,4 m<sup>2</sup>/ha de área basal. Se seleccionaron aleatoriamente 115 árboles muestra distribuidos en diferentes unidades muestrales dentro de los rodales. A cada uno de los árboles se les midió diámetro a los 1,3 metros ( $d$ ), altura total ( $h$ ), altura de fuste (definida como la altura de fin de aprovechamiento maderable del árbol, diámetro de utilización de 7,5 cm), y la altura de cada una de las secciones de corte (el largo de las secciones variaron entre 1,5 a 2,5 m). En cada sección se extrajo la rodaja final (4 cm de espesor), para posteriormente realización en laboratorio de análisis fustal o de tronco. En el presente estudio, del análisis fustal tan sólo se empleó la medición del diámetro de cada

rodaja. Posteriormente se cubió cada sección mediante la fórmula del tronco de cono (Prodan *et al.*, 1997, Husch *et al.*, 2003), con lo cual se obtuvo el volumen maderable con corteza. El diámetro fustal a los 4 m de altura sobre el fuste ( $d_4$ ) fue interpolado cúbicamente en cada árbol. Posteriormente, la muestra fue aleatoriamente dividida en 85 árboles empleados para el proceso de ajuste y 30 árboles empleados para la validación independiente de los modelos (Tabla 1).

**Tabla 1.** Estadística descriptiva de los árboles muestra.

Estadístico	Muestra de ajuste ( $n=85$ )				Muestra de validación ( $n=30$ )			
	$d$ (cm)	$h$ (m)	$d_4$ (cm)	$V$ ( $m^3 scc$ )	$D$ (cm)	$h$ (m)	$d_4$ (cm)	$v$ ( $m^3 scc$ )
Mínimo	10,0	6,0	5,9	0,021	12,5	8,7	7,7	0,055
Máximo	57,0	34,3	22,7	1,454	42,0	25,6	17,9	0,813
Media	27,1	17,7	13,4	0,372	27,09	17,8	12,9	0,328
CV(%)	36,7	38,3	24,6	89,2	26,2	22,5	19,2	52,2

El volumen es la variable que presenta la mayor variabilidad (Tabla 1), lo cual es concordante con otros estudios (Salas, 2002). La variación del volumen detectada en el estudio es mayor a la reportada por Diéguez *et al.* (2003), quienes midieron una mayor cantidad de árboles. Esto indica que la muestra empleada cubrió diversos tipos y gestiones de montes en el área de estudio.

## 2.2. Modelos analizados

Utilizando un programa informático escrito en MATLAB versión 6.5.1, (The MatWorks Inc., 2002) se ajustaron mediante el método de los mínimos cuadrados ordinarios y compararon los modelos estimadores de volumen especificados en la tabla 2. Dichos modelos constituyen la mayor parte de los tradicionales modelos de volumen empleados en diversos estudios (Spurr, 1952; Prodan *et al.*, 1997; Salas, 2002; Husch *et al.*, 2003).

**Tabla 2.** Modelos de volumen ajustados

Número	Modelo	Fuente o denominación
1	$v = \beta_0 + \beta_1 d$	Lineal simple
2	$v = \beta_0 + \beta_1 d^2$	Línea de volumen
3	$v = \beta_0 + \beta_1 d^3$	
4	$v = \beta_0 + \beta_1 d + \beta_2 d^2$	Cuadrático
5	$v = \beta_0 + \beta_1 d + \beta_2 d^{0.5}$	Manosalva (1995)
6	$\ln(v) = \beta_0 + \beta_1 \ln(d)$	Meyer (1953)
7	$\ln(v) = \beta_0 + \beta_1 \ln(d) + \beta_2 / \ln(d)$	Manosalva (1995)
8	$\ln(v) = \beta_0 + \beta_1 \ln(d) + \beta_2 / \ln(d^2)$	Manosalva (1995)
9	$v = \beta_0 + \beta_1 d^2 h$	Spurr (1952)
10	$\ln(v) = \beta_0 + \beta_1 \ln(d^2 h)$	Logarítmico de Spurr (1952)
11	$\ln(v) = \beta_0 + \beta_1 \ln(d) + \beta_2 \ln(h)$	Schumacher y Hall (1933)
12	$v = \beta_0 + \beta_1 d d_4 h$	Zöhler (1980)
13	$v = \beta_0 + \beta_1 d^2 + \beta_2 d^2 h^2 + \beta_3 d_4^2$	Salas (2002)
14	$v = \beta_0 + \beta_1 d^2 h + \beta_2 d^2 h^2 + \beta_3 h^3 + \beta_4 d_4^2$	Salas (2002)
15	$d^2 / v = \beta_0 + \beta_1 \ln(h) + \beta_2 \ln(d_4) + \beta_3 d^2 h$	Salas (2002)
16	$\ln(v) = \beta_0 + \beta_1 \ln(d) + \beta_2 \ln(h) + \beta_3 \ln(d_4)$	Zöhler (1980)
17	$\ln(v) = \beta_0 + \beta_1 \ln(d_4) + \beta_2 \ln(h) + \beta_3 d^2 h$	Salas (2002)

Donde:  $v$  = volumen maderable ( $m^3 scc$ );  
 $d$  = diámetro con corteza a la altura del pecho (cm);  
 $h$  = Altura total (m);  
 $d_4$  = Diámetro con corteza del fuste a los 4 metros de altura en el árbol (cm).

Los modelos pueden agruparse de acuerdo a las variables forestales que son necesarias de medir en el campo para emplear la ecuación, en a) de simple entrada, requieren conocer el  $d$  del árbol; b) de doble entrada, los que además del  $d$  emplean  $h$ ; y c) de triple entrada, en donde se requiere la medición de un diámetro fustal medido a una altura superior a la altura del diámetro normal, denominado  $d_z$ . Estos últimos tipos de modelos han mostrado un muy buen comportamiento en la estimación de volumen de diversas especies (Parresol y Hotvedt, 1990; Navarro, 1995; van Laar y Akça, 1997; Salas, 2002). Es necesario aclarar que en los modelos originales en los que aparece un valor del diámetro en altura se han utilizado tanto el diámetro a la altura de 3 m (Núñez y Salas, 2000), a los 4,5 m (Rodríguez, 1995; Martínez *et al.*, 1993), así como también a los 6 m (Salas, 2002). En el presente estudio, para poder comparar resultados con las ecuaciones propuestas por los inventarios forestales nacionales (IFN) españoles, se ha empleado el diámetro a 4 metros que utiliza dicho inventario.

### 2.3. Selección del modelo

La selección de la ecuación de volumen se realizó de acuerdo a la metodología propuesta por Salas (2002), la cual consiste en las siguientes cuatro etapas:

a) Cumplimiento de supuestos. Se verificó el cumplimiento de los supuestos del análisis de regresión. La normalidad, homocedasticidad, e independencia serial de los residuales se verificó mediante el test de Kolmogorov-Smirnov (Ostle, 1973), prueba de Levene (Ott y Longnecker 2001), y test de Durbin y Watson (Canavos, 1988), respectivamente. La presencia de multicolinealidad entre las variables predictoras, fue evaluada mediante el factor de inflación de la varianza (*VIF*) (Maddala, 1996). Los modelos que no cumplieron con algunos de estos supuestos fueron eliminados de las siguientes etapas. La presencia de multicolinealidad se asumió para aquellas ecuaciones que presentaron *VIF* mayores a 10 para alguna de sus variables independientes (Kozak, 1997).

b) Bondad de ajuste. Se calculó el coeficiente de determinación ajustado ( $R^2_{adj}$ ), el error estándar de la estimación ( $S$ ) y la significancia estadística de los coeficientes de regresión ajustados mediante la prueba de *t*-student. Aquellos modelos que presentaron al menos un parámetro no significativamente distinto de cero fueron eliminados. Para comparar el error estándar de la estimación (comúnmente calculado en análisis de regresión) de ecuaciones con distintas variables dependientes, se calculó el índice de Furnival “*IF*” (Furnival, 1961). Este índice se obtiene maximizando la función de verosimilitud de cada modelo de volumen, en el espacio muestral de la variable dependiente  $v$ . La interpretación del *IF* es de carácter inverso a los de máxima verosimilitud, por lo tanto un valor grande indica ajuste pobre y viceversa (Furnival, 1961).

c) Capacidades predictivas. Fueron evaluadas en la totalidad de la muestra empleada para la validación de los modelos (validación independiente) las capacidades predictivas, con el fin de comparar los modelos. Los estadísticos empleados son la raíz del error cuadrático medio (*RECM*) y la diferencia agregada (*DA*) (Prodan *et al.*, 1997). Debido al sesgo que se produce al transformar los valores estimados de modelos logarítmicos ( $\ln m^3$ ) a los valores aritméticos ( $m^3$ ), se empleó (previo al cálculo del *RECM* y *DA*) la corrección propuesta por Baskerville (1972) para los modelos que emplean logaritmos en la variable respuesta.

d) Análisis de capacidades predictivas dentro de clases de validación. Se calcularon la media y la desviación estándar del *RECM* y *DA* calculados para cada rango de validación de cada ecuación (Salas, 2002). Se emplearon cinco rangos de validación de acuerdo al volumen observado ( $v$ ) de cada árbol. Los rangos fueron establecidos de acuerdo a los percentiles 20, 40, 60 y 80, con lo cual cada uno de los cinco rangos tuvo la misma cantidad de observaciones.

En general, con esta metodología, se espera que un modelo con un buen comportamiento debiera presentar para el *RECM* una menor media (mayor exactitud) y una menor desviación

estándar (menos variabilidad) en todos los rangos. Para la *DA*, la tendencia esperada es la misma, con la diferencia que para la obtención de la media y desviación estándar de la *DA* en las clases de validación se emplean valores absolutos. Este análisis es clasificado de acuerdo a Kozak y Kozak (2003) como un análisis de falta de ajuste, el cual normalmente no es realizado en gran parte de los estudios que evalúan y comparan varios modelos.

La elección de la mejor ecuación estimadora de volumen, se efectuó a través de la asignación de puntajes a cada uno de los estadísticos (Cao *et al.*, 1980; Salas, 2002). El valor 1 fue asignado a la correspondiente ecuación ajustada que presentaba; a) el menor valor de *RECM* calculado para toda la muestra, b) el menor valor de la media y c) desviación estándar de la *RECM* en las clases de validación hasta completar la asignación de puntajes con el total de modelos. De igual manera se realizó con el análisis del sesgo, esto es, el valor 1 fue asignado a la ecuación que presentó d) la *DA* más cercana a cero calculada para toda la muestra, e) el menor valor de la media en valor absoluto y f) desviación estándar de la *DA* en las clases de validación. Finalmente, se obtuvo la sumatoria de estos puntajes para cada modelo, donde los menores puntajes finales indican ecuaciones más exactas. En caso que hubiera ecuaciones con el mismo puntaje se optó por la expresión matemática más sencilla.

## 2.4. Comparación con otras ecuaciones publicadas

Para verificar la importancia del estudio, se compararon los valores estimados por la ecuación seleccionada, versus los valores estimados mediante los modelos propuestos para *Pinus pinaster* para Pontevedra de los inventarios forestales nacionales (IFN) españoles y el de Rodríguez (1995) sobre una muestra independiente y no empleada en el proceso de ajuste de nuestros modelos. Se utilizaron dos ecuaciones, la primera de ellas emplea como variables predictoras el diámetro normal (*d*) y la altura total (*h*), dado que estas son variables que normalmente se registran en inventarios forestales, se ha denominado “modelo general del IFN”. La expresión matemática es la siguiente (MINISTERIO DE MEDIO AMBIENTE, 2002);

$$v = 41,10 + 0,0002834(d^2h), \quad [1]$$

donde, las variables corresponden a las previamente definidas, pero con diferentes unidades de medición (*v* es expresado  $\text{dm}^3$  y *d* en mm), las cuales obviamente fueron transformadas para realizar las comparaciones. El modelo de Rodríguez (1995) emplea las mismas variables (aunque *d* es expresado en cm), y es propuesto para la zona costera de Galicia,

$$v = 0,0755567(d^2) + 0,030658(d^2h). \quad [2]$$

La tercera ecuación a comparar con la propuesta en este artículo, requiere además de las variables tradicionales, la medición de un diámetro fustal en altura, por lo cual la hemos denominado “modelo con diámetro superior del IFN”. La expresión matemática es la siguiente (MINISTERIO DE MEDIO AMBIENTE, 1993);

$$v = 0,0022790 d^{1,59308} h^{0,63260} d_4^{0,44296}, \quad [3]$$

donde, las variables corresponden a las definidas anteriormente para la ecuación simple y la variable  $d_4$  esta expresada en cm.

## 3. RESULTADOS Y DISCUSION

### 3.1. Selección de la ecuación de volumen

En términos generales los modelos de volumen de simple y doble entrada presentaron en promedio un *RECM* del 20,69% y 17,67% respectivamente, los cuales son notoriamente mayores a los modelos de volumen de triple entrada, los que tan sólo tienen en promedio un *RECM* de 9,79%. Similares tendencias han sido encontradas en otros estudios (Parresol y Hotvedt, 1990; Navarro, 1995). Debido a las notables diferencias en exactitud en los tipos de

modelos, sólo se presentan en este artículo los cuatro mejores modelos, los cuales, dado lo anterior, obviamente poseen alguna expresión del diámetro superior ( $d_4$ ) en su formulación.

a) *Cumplimiento de supuestos*. Los modelos logarítmicos logran normalizar la distribución de frecuencias y la homogeneidad de la varianza de los residuales (Tabla 3). El uso del logaritmo ha sido ampliamente recomendado en la modelación del volumen (Prodan *et al.*, 1997; van Laar y Akça, 1997), sobre todo cuando se están analizando datos con amplio rango de ajuste, en donde el efecto de la heterocedasticidad es más notorio (Salas, 2002).

**Tabla 3.** Parámetros estimados y su correspondiente significancia, y estadísticos empleados para la verificación de supuestos

Modelo		Parámetros estimados				Supuestos estadísticos			
		$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	Norm.*	Homoce.†	Autocor.††
12	valor	0,0268	3,663E-05				1,489	9,753	1,809
	v-p	0,004	0,000				0,024	0,000	
	VIF		1,000						
14	valor	-0,0147	-1,55E-05	7,71E-07	3,13E-06	0,0010	0,900	9,016	1,941
	v-p	0,000	0,341	0,422	0,000	0,000	0,393	0,000	
	VIF		102,129	78,711	5,049	8,889			
16	valor	-87,170	0,1407	0,6607	0,8938		0,563	0,428	1,711
	v-p	0,000	0,000	0,000	0,000		0,910	0,734	
	VIF		11,384	2,044	10,534				
17	valor	-83,837	18,491	0,6257	2,61E-03		0,568	0,613	1,793
	v-p	0,000	0,000	0,000	0,043		0,904	0,608	
	VIF		2,928	2,224	3,180				

\* Normalidad. Valor corresponde al estadístico de bondad de ajuste de Kolmogorov-Smirnov, y su respectiva significancia estadística (v-p).

† Homocedasticidad. Valor corresponde al estadístico de Levene, y su respectiva significancia estadística (v-p).

†† Autocorrelación. Valor corresponde al estadístico de Durbin-Watson.

Aunque teóricamente la independencia serial de primer orden de los residuos puede ser mayormente quebrantada cuando se ajustan modelos de ahusamiento (Kozak, 1997) y de crecimiento (Monserud, 1986), el modelo 11 presenta autocorrelación en sus residuos, al comparar los valores del test de Durbin y Watson con los tabulares publicados por Canavos (1988) de acuerdo al número de parámetros y usando un nivel de confianza estadística del 95%. Esto debiera estar explicado, ya que algunos árboles muestra se obtuvieron desde las mismas unidades muestrales, por lo tanto, entre ellos podría existir una correlación espacial que este influenciando los resultados del modelo. No obstante, empleando los otros modelos (12, 14, 17) la posible correlación serial es eliminada, o bien no detectada por el test de Durbin y Watson. Mayores análisis podrían ser conducidos al respecto, por ejemplo empleando un modelo autoregresivo, no obstante aquello, fue considerado no necesario en este estudio.

El modelo 14 presenta multicolinealidad severa (Tabla 3), ya que presenta variables predictoras independientes correlacionadas entre si y redundan sobre el modelado de la variable respuesta (Salas, 2002). El modelo 16 entre tanto también presenta multicolinealidad, aunque algo más moderada. Tan sólo el modelo 17 (ya que el modelo 12 tiene un sólo termino como variable predictor) no evidenció multicolinealidad entre sus variables predictoras, lo cual se debe a que las variables son representadas en la expresión no tan linealmente como en el modelo 16. La multicolinealidad afecta en forma severa las estimaciones de mínimos cuadrados, ya que éstas tienden a ser menos precisas para los efectos individuales de las variables independientes, es decir, cuando dos o más variables de predicción son colineales los parámetros no miden los efectos individuales sobre la respuesta, sino que reflejan un efecto parcial sobre la misma, sujeto

a todo lo que pase con las demás variables de predicción en la ecuación de regresión (Canavos, 1988).

En términos generales, el no cumplimiento de los supuestos de normalidad, homocedasticidad, e independencia serial de los residuos, así como la independencia de las variables predictoras, tiene diversos efectos en el método de mínimos cuadrados ordinarios, los cuales básicamente tienen relación a que; las pruebas de significancia (hipótesis) no son validadas, los estimadores son ineficientes (aunque siguen siendo insesgados), y los estimadores de las varianzas son sesgados (Furnival, 1961; Baskerville, 1972; Ostle, 1973; Canavos, 1988, Maddala, 1996). Además, dado que el objetivo final del estudio es desarrollar modelos que sean aplicados a la población probable de árboles de *Pinus pinaster* en la región, es importante contar con las mayores confianzas probabilísticas y de inferencia posibles de los parámetros ajustados en los modelos, por lo que el cumplimiento de los supuestos que sustentan el análisis de regresión son vitales, aún más con una muestra de datos reducida. Aunque el único modelo que supero la etapa de cumplimiento de supuestos estadísticos, es el 17, de todas maneras, se seguirá comparando con los otros tres modelos para ilustrar los resultados obtenidos.

b) *Bondad de ajuste*. La significancia estadística de todos los parámetros estimados (Tabla 3), fue solamente cumplida por los modelos 12, 16 y 17. De acuerdo a la asignación de puntajes en base al *IF*, del cual se obtuvo el ranking de la Tabla 4, los mejores modelos son el 17 y el 16. Si bien el coeficiente de determinación fue similar para estos modelos, existe una diferencia de un 100% menos en el *IF* para los modelos logarítmicos versus no logarítmicos. Estas diferencias fueron también indicadas por Furnival (1961), y se pueden deber a la sensibilidad del índice al emplear derivadas.

**Tabla 4.** Estadísticos y ranking de modelos de acuerdo a la bondad de ajuste y validación.

Modelo (número)	<i>S</i> (m <sup>3</sup> scc)	Bondad de ajuste			Validación				Ranking <sup>†</sup>
		$R^2_{adj}$	<i>IF</i> (m <sup>3</sup> scc)	Ranking <sup>*</sup>	<i>RECM</i> (m <sup>3</sup> scc)	(%)	<i>DA</i> (m <sup>3</sup> scc)	(%)	
12	0,0670	0,9699	0,0670	4°	0,03660	9,332	-0,0001	-0,0382	3°
14	0,0425	0,9879	0,0425	3°	0,03835	9,777	-0,0050	-1,2685	4°
16	0,0743	0,9808	0,0228	2°	0,03050	7,768	-0,0027	-0,4599	2°
17	0,0728	0,9811	0,0224	1°	0,03048	7,725	0,0010	0,2526	1°

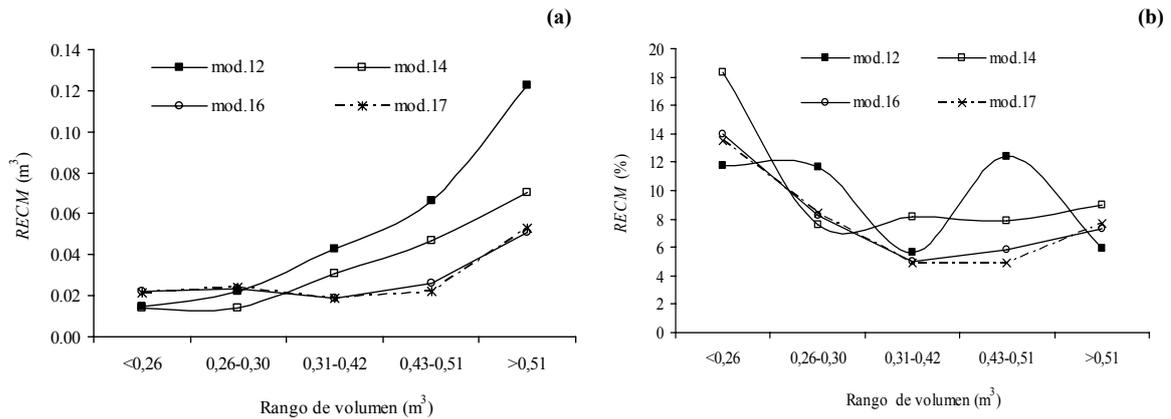
\* Basado en el índice de Furnival (1961).

† Basado en el *RECM* y *DA* en las unidades del volumen, tanto para toda la muestra como también del análisis por clases de validación.

c) *Capacidades predictivas*. La exactitud global (no segregando por clases de validación) de los modelos es considerada satisfactoria, ya que ninguna ecuación supero el 10% en *RECM* (Tabla 4), mostrando un buen comportamiento predictivo. Las diferencias en la medida de la exactitud a nivel global son mínimas.

Los modelos son insesgados, aunque el modelo 14 supero el 1% de *DA* (Tabla 4). Aunque, cuando se usa el método de mínimos cuadrados ordinarios, por definición el sesgo promedio debe ser igual a cero (Kozak y Kozak, 2003), los valores porcentuales del sesgo aquí reportados (levemente distintos a cero), se pueden deber básicamente a que el factor de corrección de Baskerville (1976), usado para remover el sesgo producido al transformar los valores logarítmicos a aritméticos, no fue lo suficiente fino para remover completamente el sesgo. El modelo 16 sobreestima dos veces más que el modelo 17. Además, el modelo 17 ofrece la ventaja práctica en toma de decisiones en manejo forestal, de una leve tendencia a subestimar.

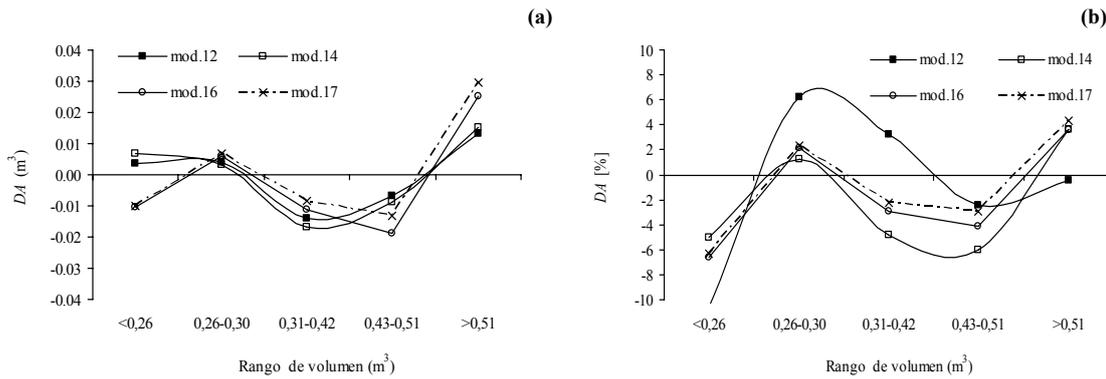
d) *Análisis de las capacidades predictivas dentro de rangos de clases volumétricas*. La exactitud de los modelos disminuye a medida que el diámetro normal de los árboles aumenta (Figura 1a). Esto se debe básicamente a que los árboles presentan una mayor dispersión en su volumen a medida que su tamaño aumenta, lo que provoca heterocedasticidad de la varianza residual.



**Figura 1.** Estadístico de exactitud (*RECM*) en volumen (a) y porcentaje (b) para los modelos por clases de validación.

Las mayores diferencias de exactitud se dan entre los modelos no logarítmicos (12 y 14), versus los modelos logarítmicos (17 y 20). Los modelos 12 y 14 además de poseer menores exactitudes, presentan las mayores variaciones en sus estimaciones, los modelos 16 y 17, en cambio, son los que presentan menor variabilidad en sus *RECM* entre clases de validación. Entre estos dos últimos modelos las mayores diferencias de exactitud se dan en la cuarta clase de validación (rango de volumen), y muestran en general una homogeneidad en la exactitud hasta la tercera clase. La figura 1b, muestra la *RECM* en términos porcentuales (dividiendo el estadístico por el promedio del volumen observado en cada clase), mostrando esta vez una tendencia inversa, es decir, estas ecuaciones poseen mayores errores porcentuales a menores diámetros, y una notoria variación en la exactitud del modelo 12. No obstante, los errores para los árboles que concentran los mayores volúmenes, no superan el 8% con los mejores modelos. Los errores porcentuales, de alguna manera encubren la verdadera magnitud de la falta de exactitud, (debido a que así como los errores son mayores para los árboles de grandes dimensiones, éstos también poseen los mayores volúmenes, por lo tanto al calcular el error porcentual, la tendencia a la heterocedasticidad es disminuida) pero son de gran utilidad práctica para el lector que no está familiarizado con las unidades de medición empleadas, ni los rangos de volumen que puede alcanzar la especie en estudio.

El sesgo de los modelos presenta una tendencia a la subestimación en las clases superiores y sobreestimación para árboles pequeños, y una clara sobreestimación para los volúmenes medios (Figura 2a). Esta sobreestimación probablemente puede deberse a una mayor variabilidad en dichas clases, en donde alguna(s) observación(es) de gran(es) volumen(es) para esa clase en particular, pudiera(n) forzar (debido al ajuste de mínimos cuadrados) a los modelos a estimar valores cercanos a esas observaciones, pero por encima de la gran parte de los otros árboles que forman parte de esa clase. La representación porcentual del sesgo (Figura 2b) destaca que ambos modelos sobreestiman los volúmenes en alrededor de un 6% para árboles pequeños y subestiman en un 4% en los árboles de mayores envergaduras.



**Figura 2.** Estadístico de sesgo (DA) en volumen (a) y porcentaje (b) para los modelos por clases de validación.

A igual que en el análisis de exactitud por clases, las mayores diferencias en sesgo se dan entre los modelos no logarítmicos (Figura 2). El modelo 16 tiene una leve tendencia a la sobreestimación en las clases medias de validación, en comparación con el modelo 17. Considerando que en esas clases se concentran la mayor parte de los individuos del rodal y con un importante volumen, es preferible una tendencia a la subestimación debido a las implicancias económicas de sesgos positivos.

Luego de asignar puntajes a los valores medios y desviaciones estándar de los estadísticos de validación de cada modelo, el modelo 17 (Salas, 2002) presentó la mejor evaluación (Tabla 4) debido a una mayor estabilidad en sus estimaciones en las clases de validación, así como en sus estadísticos de validación global. Aunque los contrastes entre los modelos 16 y 17 no son tan notorios, la metodología de selección empleada, permite detectar pequeñas diferencias entre las ecuaciones. Además, se prefiere el uso del modelo 17, ya que es el único que cumplió con todos los supuestos estadísticos.

Ahora bien, ya que el modelo 17 fue corregido por el factor de Baskerville (1972), la mitad de la varianza de la ecuación debe ser agregada al intercepto del modelo. La ecuación final y propuesta para estimar el volumen maderable con corteza de *Pinus pinaster* en el Baixo Miño es:

$$\ln(v) = -8,3811 + 1,8491 \ln(d_i) + 0,6257 \ln(h) + 0,000002606 d^2 h \quad [4]$$

### 3.2. Comparación con otras ecuaciones publicadas

El resultado obtenido al comparar la ecuación seleccionada en este estudio con las estimaciones realizadas con las ecuaciones propuestas por los inventarios forestales nacionales para *Pinus pinaster* en Pontevedra (expresiones 1 y 2) y el modelo de zona costera de Rodríguez (1995) (expresión 3), se dan en la tabla 5.

**Tabla 5.** Estadísticos de exactitud y sesgo para los modelos del inventario forestal nacional y el propuesto en este estudio

Denominación del modelo	Estadístico			
	RECM		DA	
	(m³scc)	(%)	(m³scc)	(%)
IFN general	0,07171	18,29	-0,0261	-6,64
Rodríguez (1995)	0,12519	31,92	-0,0749	-19,11
IFN diámetro superior	0,06270	15,98	-0,0243	-6,18
Propuesto	0,03048	7,77	0,0001	0,25

La ecuación del IFN que requiere el mismo número de variables de entrada (modelo con diámetro superior) que la ecuación propuesta, entrega estimaciones sesgadas, sobreestimando el volumen en un 6,64%, respectivamente. El error (en porcentaje) que se cometería al emplear dicha ecuación del IFN en los rodales en estudio, es un 105,5% mayor que el obtenido al emplear el modelo propuesto. Al hacer las comparaciones con modelos que tan solo empleen el  $d$  y  $h$  como predictores, el error como era de esperar, aumentan a un 18% y el sesgo a un 6.64% de sobreestimación para el modelo del IFN, y tanto el error como el sesgo son considerablemente superiores al emplear el modelo de Rodríguez (1995). Esta tendencia es lógica pues la expresión 1 y 2, sólo explican el volumen mediante dos variables dendrométricas, mientras que la presencia de un diámetro en altura en los modelos de volumen ha sido probada en diversos estudios satisfactoriamente. Sin embargo, las ganancias en exactitud (diferencia entre *RECM*) que se obtienen al emplear tres variables predictoras ( $d$ ,  $h$ ,  $d_d$ ) versus emplear dos ( $d$ ,  $h$ ), son solamente de un 2,31% en las estimaciones realizadas con los modelos del IFN, en comparación con las ecuaciones del presente estudio, en donde la ganancia de exactitud es de 10,51%.

Aunque si bien el modelo propuesto fue ajustado con una base de datos distinta a la cual se empleó para esta comparación con las ecuaciones del IFN, era de esperarse diferencias, debido a que tanto la muestra de ajuste y así como la de validación pertenecen a una misma región geográfica. Sin embargo, las diferencias son sustanciales. Sobre todo llama la atención las grandes diferencias que se obtienen al emplear el modelo de Rodríguez (1995), lo cual podría ser un indicio de que la división geográfica de las ecuaciones del IFN es más específica que la planteada por Rodríguez (1995). Este hecho justifica el desarrollo de modelos dendrométricos específicos para ciertas comarcas naturales (eco-regiones) en Galicia, en vez de la reciente tendencia a desarrollar modelos generales para Galicia (Rodríguez 1995, Rojo *et al.*, 2005).

Otro factor que influye estas diferencias, son los rangos de ajuste de los modelos. La muestra empleada por Rodríguez (1995) en el ajuste de su modelo, aunque notoriamente mayor (678 árboles en costa), cubría tan solo hasta árboles con 37 cm de diámetro normal y 26 m en altura, y nuestra muestra cubre hasta los 57 cm ( $d$ ) y 34.3 m ( $h$ ), por lo tanto al emplear el modelo de Rodríguez (1995), se están obteniendo estimaciones para un rango de las variables predictoras que no fue parte del modelo original. Esto además, es una clara indicación del riesgo que supone la extrapolación de modelos de regresión fuera de los rangos de ajuste. Por otra parte, no fue posible conocer los rangos de ajuste empleadas por el IFN, así como tampoco alguna medida de exactitud y sesgo originalmente calculada para las ecuaciones, ya que no fueron publicados en los respectivos documentos. No obstante, las diferencias encontradas en las capacidades predictivas de las ecuaciones, son notables, aumentando en promedio la *RECM* y la *DA* del modelo con diámetro superior del IFN en un 105,5% y 2349% respectivamente, en comparación con la ecuación seleccionada y propuestas en el presente estudio.

Estas diferencias cobran una importancia vital, debido a que en los procesos normales de inventarios forestales, cada volumen individual será agregado a nivel de rodal, lo cual provoca que las posibles estimaciones de volumen comercial en pie y por ende, sus tasaciones incurrirán en un fuerte error, que se traducirá en pérdidas monetarias.

En definitiva, el modelo propuesto implica una ganancia en exactitud y disminuye el sesgo en las estimaciones de volumen, que se realizarían empleando otros modelos de volumen actualmente en uso. Una mejor aproximación al tema (estimación de volumen individual de *Pinus pinaster* en el Baixo Miño) podría resolverse empleando una mayor y amplia base muestral, por ejemplo, empleando los árboles muestras empleados en el tercer y segundo inventario forestal, así como también la muestra empleada por Rodríguez (1995), Diéguez *et al.* (2003), y Rojo *et al.* (2005).

#### 4. CONCLUSIONES

La inclusión de un diámetro superior (medido en altura) en los modelos de volumen resulta vital, para lograr buenas estimaciones, sin embargo, en los procesos de inventario se debería agregar su medición en una submuestra de las parcelas de muestreo, tal como se realiza con los modelos de altura.

El modelo propuesto cumple con todos los supuestos estadísticos, y además entrega las mejores predicciones.

El índice de Furnival (1961) mostró ser una buena medida para seleccionar modelos de regresión, ya que el mejor índice obtenido correspondió al modelo finalmente recomendado. Sin embargo, de todas maneras es recomendable aplicar el procedimiento de selección de ecuaciones propuesto, ya que discrimina muy bien entre ecuaciones que presentan estadísticos de validación global similares.

La aplicación de la ecuación propuesta, produce una estimación notablemente mejor que las obtenidas por otras ecuaciones propuestas para la zona costera de Galicia y la provincia de Pontevedra.

La aplicación de las ecuaciones seleccionadas en otras comarcas naturales (diferentes a la empleada para el ajuste), debe pasar por una etapa de validación, en la cual se pueda calibrar las estimaciones que éstas entregan, a través de procedimientos estadísticos, en un número reducido de árboles de prueba.

A la espera de mayor cantidad de observaciones de comarcas naturales específicas en Galicia, mayores esfuerzos podrían ser conducidos para unir las bases de datos empleadas por diversos investigadores de modelación de volumen en *Pinus pinaster* en Galicia, con el fin de desarrollar modelos de más amplia escala geográfica, empleando además una gran base muestral, pero considerando la segregación de los modelos de acuerdo a eco-regiones.

#### 5. BIBLIOGRAFIA

- Baskerville, G.L. 1972. Use of logarithmic regression in the estimation of plant biomass. *Canadian Journal of Forest Research* 2: 49-53.
- Canavos, G. 1988. Probabilidad y Estadística; Aplicaciones y Métodos. Trad. por E. Urbina. Editorial McGraw-Hill. Primera Edición. México D.F., México. 651 p.
- Cao Q., H. Burkhart and T. Max. 1980. Evaluation of two methods for cubic-volume prediction of Loblolly Pine to any merchantable limit. *Forest Science* 26 (1): 71-80.
- Diéguez, U., M. Barrio, F. Castedo y M. Balboa. 2003. Estimación del diámetro normal y del volumen del tronco a partir de las dimensiones del tocón para seis especies forestales comerciales de Galicia. *Investigación agraria. Sistemas y Recursos Forestales* 12 (2): 131-139.
- Husch, B., T. W. Beers and J.A. Kershaw. 2003. *Forest Mensuration*. Fourth edition. Wiley. New York, USA. 443 p.
- Furnival, G. 1961. An index for comparing equations used in constructing volume tables. *Forest Science* 7 (4): 337-341.
- García, O. 2001. Aproximación a la dinámica de los montes gallegos. Existencias y evolución de los tipos forestales en la Provincia de Pontevedra. *Investigación agraria. Sistemas y Recursos Forestales* 10 (2): 291-310.
- Kozak, A. 1997. Effects of multicollinearity and autocorrelation on the variable-exponent taper equations. *Canadian Journal of Forest Research* 27: 619-629.
- Kozak, A. and R. Kozak. 2003. Does cross validation provide additional information in the evaluation of regression models?. *Canadian Journal of Forest Research* 33: 976-987.
- Maddala, G. 1996. *Introducción a la Econometría*. Trad. por J. Jolly. Prentice-Hall Hispanoamericana, S.A. Segunda Edición. México D.F., México. 715 p.
- Manosalva, L. 1995. Antecedentes dendrométricos básicos de un rodal de Lenga (*Nothofagus pumilio* (Poepp. et Endl.) Krasser), en el sector Cerro La Virgen, XI Región. Tesis Ingeniero Forestal. Universidad Austral de Chile. Valdivia, Chile. 69 p.
- Martínez, J., P. Ara y F. González. 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(2): 211-228.

- Meyer, H. A. 1953. Forest Mensuration. Penn's Valley Publishers, State College, PA.
- MINISTERIO DE MEDIO AMBIENTE. 1993. Segundo inventario forestal nacional 1986- 1995. Galicia – Pontevedra. Instituto para la Conservación de la Naturaleza. ICONA. España. 233 p.
- MINISTERIO DE MEDIO AMBIENTE. 2002. Tercer inventario forestal nacional 1997- 2006. Galicia – Pontevedra. Dirección General de la Conservación de la Naturaleza. España. 314 p.
- Monserud, R. A. 1986. Time-series analyses of tree-ring chronology. *Forest Science* 32(2): 349-372.
- Navarro, R. 1995. Ecuaciones de volumen con diámetros en altura. Tesis Ingeniero Forestal. Universidad Austral de Chile. Valdivia, Chile. 29 p.
- Núñez, P. y Salas C. 2000. Estudio dendrométrico proyecto forestal Río Cóndor, Tierra del Fuego, XII Región, Chile. Forestal Savia Ltda. Documento de circulación restringida. Temuco, Chile. 43 p.
- Ostle, B. 1973. Estadística Aplicada; Técnicas de la Estadística Moderna, Cuando y Donde aplicarlas. Trad. por D. De La Serna. Editorial Limusa - Wiley S.A. Tercera Reimpresión. México D.F., México. 629 p.
- Ott, R.L. and M. Longnecker. 2001. An introduction to statistical methods and data analysis. Fifth Edition. Duxbury. Pacific Grove, USA. 1152 p.
- Parresol, B. R. and J. E. Hotvedt. 1990. Diameter measurement in bald cypress. *Forest Ecology and Management* 33/34: 509-515.
- Prodan, M., R. Peters, F. Cox, y P. Real. 1997. Mensura Forestal. Serie Investigación y Educación de Desarrollo Sostenible. Instituto Interamericano de Cooperación para la Agricultura (IICA)/BMZ/GTZ. San José, Costa Rica. 561 p.
- Robinson, A. P. and G. B. Wood. 1994. Individual tree volume estimation: a new look at new systems. *Journal of Forestry* 92(12): 25-29.
- Rodríguez, R. 1995. Crecimiento y producción de masas forestales regulares de *Pinus pinaster* Aiton. en Galicia. Alternativas selvícolas posibles. Tesis doctoral, ETSI de Montes, Universidad Politécnica de Madrid. Madrid, España. 297 p.
- Royo, A., X. Perales, F. Sánchez, J. G. Álvarez, and K. von Gadow. 2005. Stem taper functions for maritime pine (*Pinus pinaster* Ait.) in Galicia (Northwestern Spain). *European Journal of Forest Research* 124(3): 177-186.
- Salas, C. 2002. Ajuste y validación de ecuaciones de volumen para un relicto del bosque de Roble-Laurel-Lingue. *Bosque* 23(2): 81-92.
- Schumacher, F. X. and S. H. Hall. 1933. Logarithmic expression of timber tree volume. *Journal of Agr. Research* 47: 719-724.
- Spurr, S. H. 1952. Forest Inventory. Ronald Press. Co., New York. 416 p.
- The MathWorks Inc., 2002. MATLAB, The language of technical computing, release 13. Natick. Mass., USA.
- van Laar A. and A. Akça. 1997. Forest Mensuration. Cuvillier Verlag. Göttingen, Alemania. 418 p.
- Zöhler F., 1980. Forstinventur. Ein Leitfaden für Studium und Praxis. Pareys Studentexte 26. Verlag Paul Parey. Hamburg und Berlin, Alemania. 207 p.

