

## ARTÍCULOS

## Métodos estadísticos paramétricos y no paramétricos para predecir variables de rodal basados en Landsat ETM+: una comparación en un bosque de *Araucaria araucana* en Chile

Parametric and non-parametric statistical methods for predicting plotwise variables based on Landsat ETM+: a comparison in an *Araucaria araucana* forest in Chile

Christian Salas<sup>a,b\*</sup>, Liviu Ene<sup>c</sup>, Nelson Ojeda<sup>a</sup>, Héctor Soto<sup>a</sup>

\*Autor de correspondencia: <sup>a</sup>Universidad de La Frontera, Departamento de Ciencias Forestales, Laboratorio de Análisis Cuantitativo de Recursos Naturales, casilla 54-D,

Temuco, Chile, tel.: +56(45) 325658, fax.: +56(45) 325664, christian.salas@yale.edu

<sup>b</sup>Yale University, School of Forestry and Environmental Studies, New Haven, USA.

<sup>c</sup>Norwegian University of Life Sciences, Department of Ecology and Natural Resources Management, Ås, Norway.

## SUMMARY

The *Araucaria araucana* forests have a high level of both ecological and scientific importance, because they are long-lived and endemic. Although there have been several ecological studies conducted concerning *A. araucana* forests, none has produced quantitative models. We compared parametric and non-parametric statistical methods for predicting stand variables from Landsat ETM+ derived variables from two *A. araucana* stands in south-central Chile. The assessed parametric methods were multiple linear regressions (MLR), generalized least squares with a non-null correlation structure (GLS), linear mixed-effects models (LME), and partial least squares (PLS); while the non-parametric methods were: *k*-nearest neighbor (*k*-NN) and most similar neighbor (MSN). In descending order, number of trees per ha (*N*), stand gross volume (*V*), stand basal area (*G*), and dominant height ( $H_{dom}$ ) were the most difficult variables to be modeled by all the methods. LME with known random effects (i.e., LME1) performed best, achieving a root mean square showing differences (RMSD) for *N* and *V* of 18.31 and 4.08 % versus 33.06 and 33.05 % for the second-best method, respectively. However, within the parametric methods, LME1 cannot be used for predicting new observations with no data. After LME1, GLS performed the best; also accounting for the spatial correlation of the data. Parametric methods achieved lower errors. Furthermore, differences were greater among non-parametric than those among parametric methods, with a difference of 10-15 % between *k*-NN and MSN. Although, given our results, we favor parametric methods; we point out that non-parametric methods are also useful, and the choice between parametric and non-parametric methods depends on the ultimate objective of the study.

*Key words:* *k*-nearest neighbor estimation, most similar neighbor, partial least squares regression, remote sensing.

## RESUMEN

Los bosques de *Araucaria araucana* tienen una alta importancia ecológica y científica. Aunque existen varios estudios ecológicos llevados a cabo en bosques de *A. araucana*, muy pocos han producido modelos cuantitativos. Se compararon métodos estadísticos paramétricos y no paramétricos para predecir variables de rodal en función de variables derivadas de Landsat ETM+ para dos rodales de *A. araucana* en el centro-sur de Chile. Los métodos paramétricos fueron regresión lineal múltiple (MLR), mínimos cuadrados generalizados con una estructura de correlación no nula (GLS), modelo lineal de efectos mixtos (LME) y mínimos cuadrados parciales (PLS); mientras que los métodos no paramétricos fueron: *k*-ésimo vecino más cercano (*k*-NN) y vecino más similar (MSN). En orden descendente, número de árboles por hectárea (*N*), volumen bruto (*V*), área basal (*G*) y altura dominante ( $H_{dom}$ ), fueron las variables más complejas de modelar por todos los métodos. El modelo lineal de efectos mixtos con efectos aleatorios conocidos (LME1) tuvo el mejor desempeño, alcanzando una raíz cuadrada de las diferencias (RMSD) para *N* y *V* de 18,31 y 4,08 % versus 33,06 y 33,05 % para el segundo mejor método, respectivamente. Después de LME1, GLS se comportó mejor, y también toma en consideración la correlación espacial de los datos. Las diferencias fueron mayores entre métodos no paramétricos que para los paramétricos, con una diferencia de 10-15 % entre *k*-NN y MSN. Aunque los resultados obtenidos favorecen a los métodos paramétricos, se destaca que los métodos no paramétricos son también útiles, y la elección entre ambos métodos depende del objetivo del estudio.

*Palabras clave:* estimación del *k*-ésimo vecino más cercano, vecino más similar, regresión de mínimos cuadrados parciales, sensores remotos.

## INTRODUCCIÓN

Los bosques de *Araucaria araucana* (Mol.) K. Koch (araucaria) tienen un alto nivel de importancia ecológica. *Araucaria araucana* es una conífera que crece en el centro-sur de Chile y suroeste de Argentina (Veblen 1982). Esta especie crece en sitios altamente restrictivos, tiene tasas de crecimiento extremadamente bajas y es una especie longeva que puede alcanzar hasta 1.000 años de edad (Veblen y Schlegel 1982, González *et al.* 2006). *Araucaria araucana* está distribuida discontinuamente en las cordilleras de Los Andes y de la Costa de Chile (Veblen *et al.* 1979, 1996b, Bekessy *et al.* 2004), cubriendo 200.000 y 54.000 ha, respectivamente (CONAF/CONAMA/BIRF 1999). *Araucaria araucana* es endémica del sur de Chile y Argentina, tiene un alto valor maderero y se distribuye en una pequeña área, así como también es categorizada como una especie vulnerable (Benoit 1996). La corta de *A. araucana* en Chile ha sido prohibida desde 1976 y fue declarada Monumento Natural, así como también es incluida en el apéndice I de CITES (*Convention International Trade in Endangered Species*) (UICN 2001).

A pesar de su importancia ecológica, falta información cuantitativa de variables de rodal en bosques de *A. araucana*. Variados estudios ecológicos se han llevado a cabo en estos bosques: sobre su dinámica forestal (Montaldo 1974, Veblen 1982, Burns 1993, González *et al.* 2005), dendrocronología (Pollmann 2003, González 2005), variación genética (Bekessy *et al.* 2004), autoecología (González *et al.* 2006) y posibles tratamientos silviculturales (Puente 1980, Mujica 2001, Rechene 2003, Drake *et al.* 2005). Sin embargo, información cuantitativa de variables de rodal, tales como número de árboles, área basal, altura dominante, volumen y biomasa, no está fácilmente disponible. Esta falta de información cuantitativa ha sido también observada para otros tipos de bosques en Chile, tales como los de segundo crecimiento y adultos de roble (*Nothofagus obliqua* (Mirb.) Oerst.), raulí (*Nothofagus alpina* (Poepp *et* Endl.) Oerst.) y coigüe (*Nothofagus dombeyi* (Mirb.) Oerst.), por Salas y García (2006) y por Salas *et al.* (2006), respectivamente.

En Chile los bosques de *A. araucana* están ubicados en áreas remotas que incrementan los costos de muestreos convencionales. Para su caracterización se han aplicado técnicas de muestreo convencional, basadas en variables de árboles individuales medidas en una muestra de parcelas de terreno que son distribuidas y ubicadas de acuerdo con un diseño de muestreo predeterminado. Una característica de estos bosques en Chile es que se encuentran localizados en áreas remotas e inaccesibles, con relieve topográfico extremo, lo cual conlleva a dificultades logísticas y aumenta los costos de los muestreos convencionales. En este contexto, los sensores remotos (imágenes Landsat) permiten mejorar la eficiencia de los inventarios forestales mediante la generalización de niveles de información de parcelas sobre extensas áreas boscosas.

Las imágenes satelitales proporcionan información continua multiespectral de extensas áreas con diversas resoluciones espaciales. Cada píxel contiene un conjunto de valores enteros o números digitales (DN), que permiten cuantificar la radiancia promedio medida por el sensor en un ancho de banda específico, es decir, hay un número digital para cada canal espectral. Cada píxel puede ser considerado como un vector en un espacio espectral multidimensional. La información espectral aportada por los píxeles, denominados aquí por *X* y referidos como variables auxiliares, está disponible para toda el área de estudio. La variable de interés (la variable respuesta, representada por *Y*) es obtenida por muestreo en el terreno de una pequeña proporción de los píxeles asociados a cada parcela de terreno. En general, lo que se busca es predecir *Y* para los lugares sin mediciones de terreno, utilizando modelos y métodos estadísticos basados sobre puntos que comparten información de *Y* y *X*. Aunque las imágenes satelitales han sido ampliamente utilizadas para predecir variables de rodal en varios países (Poso *et al.* 1984, Tomppo 1991, Bauer *et al.* 1994, Trotter *et al.* 1997, Lee *et al.* 2008, entre varios otros), han sido rara vez aplicadas en bosques de Chile y Argentina (Mery y Bahamóndez 1995, Lencinas 2002, Ruiz-Tagle 2003, Bahamóndez *et al.* 2007, Lencinas y Mohr-Bell 2007, Navarro *et al.* 2008). No se están incluyendo aquí los estudios de clasificación de bosque en Chile y Argentina (Ojeda y Herrera 2000, Segura y Trincado 2003, Lizzi *et al.* 2007), porque ellos se centran en polígonos de categorías para el uso del suelo y clases de vegetación, en vez de eso, el presente estudio se centra en la predicción de variables de rodal para parcelas de muestreo.

Se han utilizado métodos estadísticos paramétricos y no paramétricos para predecir variables basadas en imágenes de sensores remotos. En la literatura relacionada, el término “imputación” se utiliza a veces en lugar de predicción, porque la imputación consiste en reemplazar las mediciones faltantes o no muestreadas para cualquier unidad de la población en estudio con mediciones de otra unidad con características similares (Ek *et al.* 1997). Así también, en estadística el término “predicción” se emplea para representar los valores estimados de la variable respuesta para datos no empleados en el ajuste de un modelo. Ya que el término predicción es más ampliamente entendido, será el utilizado en este artículo para referirse en general al empleo del modelo en estimar la variable respuesta. Existen varios métodos estadísticos<sup>1</sup> de predicción que son apropiados para estudios con sensores remotos. Entre estos métodos (o modelos) está el análisis de regresión (regresión lineal múltiple, MLR), que ofrece la base más apropiada para su aplicación en modelos univariantes (solo una variable *Y*), basado en

<sup>1</sup> Aunque en el presente artículo se ha empleado el término “métodos estadísticos”, los métodos paramétricos acá detallados son efectivamente “modelos estadísticos” y en ese contexto son a veces así referidos.

valores conocidos de  $Y$  y  $X$ . Los mínimos cuadrados generalizados (GLS) con una estructura de correlación no nula (una correlación nula se asume en MLR) toman en consideración la correlación entre observaciones con el fin de estimar correctamente los parámetros del modelo. Los modelos de efectos mixtos (ME) tienen parámetros fijos y efectos aleatorios al definir el modelo estadístico, lo cual permite acomodar la correlación de los datos. Los modelos ME han sido ampliamente utilizados en investigación forestal en años recientes para modelar variables estáticas, como ecuaciones fustales y de altura-diámetro (Leites y Robinson 2004, Mehtatalo 2004), y dinámicas, como modelos de crecimiento (Salas *et al.* 2008). Sin embargo, ME no han sido utilizados recientemente en estudios de sensores remotos, a excepción de los trabajos recientes de Meng *et al.* (2007) y Nothdurft *et al.* (2009), quienes usaron modelos lineales de efectos mixtos (LME). La regresión de mínimos cuadrados parciales (PLS) (Wold 1975, Wold *et al.* 1983) es una técnica estadística multivariante (más de una variable  $Y$ ) utilizada para predecir la respuesta de variables obtenidas mediante el uso de componentes, pero tomando en cuenta las matrices de covarianza de las variables  $Y$  y  $X$ . El  $k$ -ésimo vecino más cercano ( $k$ -NN) (Reese *et al.* 2002, Tomppo *et al.* 2008) y el vecino más similar (MSN) (Moeur y Stage 1995) son métodos no paramétricos multivariantes basados en matrices que miden la "cercanía" de algunas medidas o variables entre puntos.

La regresión lineal múltiple, los mínimos cuadrados generalizados, los modelos lineales de efectos mixtos y la regresión de mínimos cuadrados parciales son métodos paramétricos, y el  $k$ -ésimo vecino más cercano y el vecino más similar son no paramétricos. Las características matemáticas de todos los métodos antes mencionados se presentan en mayor detalle en la sección de métodos estadísticos. En comparación a los métodos paramétricos como regresión lineal múltiple, regresión de mínimos cuadrados parciales y modelos lineales de efectos mixtos, los métodos del  $k$ -ésimo vecino más cercano y el vecino más similar no hacen suposiciones de la distribución del modelo y pueden retener la estructura varianza/covarianza en la salida (Moeur y Stage 1995, Franco-Lopez *et al.* 2001, McRoberts *et al.* 2002), pero ellos no son necesariamente insesgados. Mayores antecedentes de las ventajas y desventajas de los métodos del vecino más cercano (NN:  $k$ -NN y MSN) son señalados por LeMay y Temesgen (2005).

El objetivo de este estudio es comparar métodos estadísticos paramétricos y no paramétricos para predecir variables de rodal en parcelas de muestreo usando Landsat ETM+ y mediciones de terreno en un bosque de *A. araucana* en el centro-sur de Chile. Los métodos estadísticos comparados son: regresión lineal múltiple, mínimos cuadrados generalizados, modelos lineales de efectos mixtos, regresión de mínimos cuadrados parciales,  $k$ -ésimo vecino más cercano y el vecino más similar. Según la revisión de literatura, éste

es el primer estudio aplicando variables auxiliares derivadas de imágenes satelitales para la estimación de variables de rodal de bosques de *A. araucana* en Chile, y en regiones de Argentina con clima y topografía similares.

## MÉTODOS

**Área de estudio.** Los dos rodales de *A. araucana* incluidos en el presente estudio se localizaron en el Parque Nacional Conguillío ( $38^{\circ} 38' S$ ,  $71^{\circ} 38' O$ ) en el centro-sur de Chile, a 150 km al noreste de Temuco, Chile. Conguillío cubre 60.833 ha en la Cordillera de Los Andes de la región de La Araucanía, entre los 700 y 3.125 m s.n.m. Se seleccionaron dos rodales que son representativos de las condiciones naturales de bosques de *A. araucana* en la región. El primer rodal en lo sucesivo rodal 1, cubre 274 ha, en un rango entre 1.150 y 1.380 m s.n.m.; el segundo rodal, en lo sucesivo rodal 2, cubre 71 ha, en un rango entre 1.150 y 1.225 m s.n.m. Ambos rodales se localizan en el faldeo norte del volcán Llaima y se clasifican como tipo forestal araucaria (Donoso 1981).

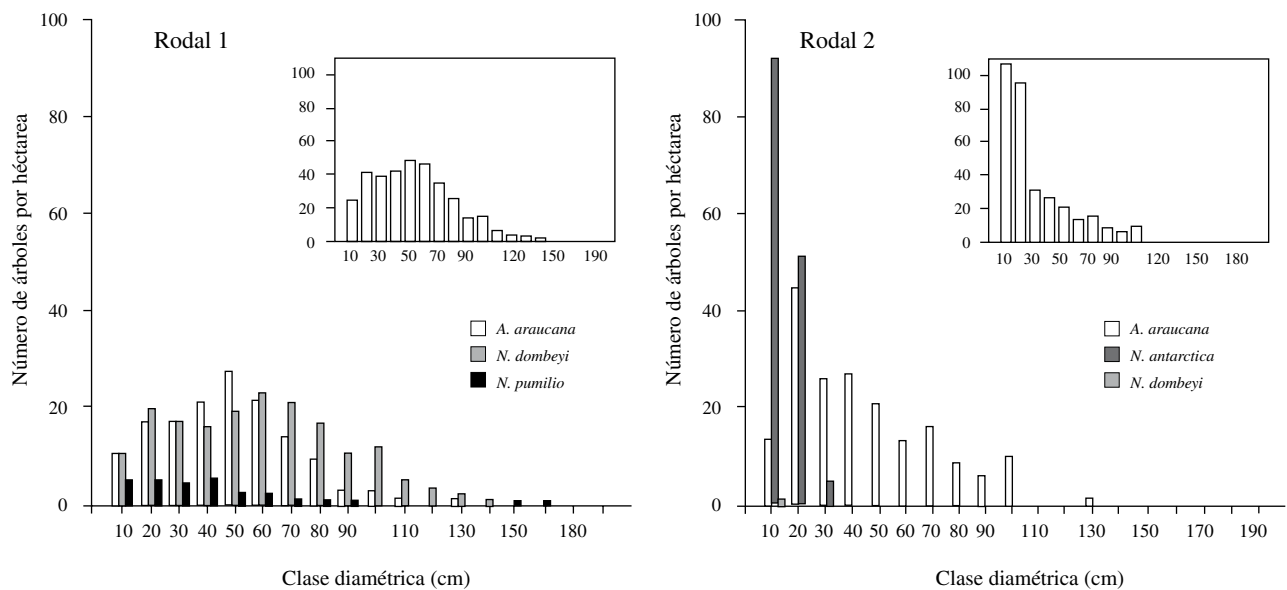
Los rodales representan distinta estructura y composición de bosques. Aunque ambos tienen densidades similares (cuadro 1), las especies dominantes en términos de área basal son *N. dombeyi*-*A. araucana* y *A. araucana*-ñirre (*Nothofagus antarctica* (G. Forster) Oerst), para los rodales 1 y 2, respectivamente. *Nothofagus dombeyi* está presente en todas las clases diamétricas del rodal 1 (figura 1) y es más abundante en el estrato del dosel superior, seguido por *A. araucana*, formando un dosel semidenso. Este tipo de bosque es representativo del tipo forestal *A. araucana* que crece entre 900 y 1.300 m s.n.m. (González *et al.* 2006). En el rodal 2, el cual tiene una menor altura dominante que el rodal 1 (cuadro 1), *A. araucana* es la especie dominante, mientras que *N. antarctica* está presente en las clases de diámetro menores (figura 1). El rodal 2 también contiene distintos parches de vegetación, debido a depósitos antiguos de rocas volcánicas. El rodal 2 se localizó en una depresión topográfica que conforma un bolsón de aire frío, un hábitat típico de *N. antarctica* (Veblen *et al.* 1996a), y conformó también una estructura común de bosques más abiertos de *A. araucana* (González *et al.* 2006).

**Datos Landsat.** Se utilizó una imagen Landsat ETM+ adquirida el 29 de noviembre del año 2001 (WRS Path233-Row 087), con una resolución de  $30 \times 30$  m ( $900 \text{ m}^2$ ). A pesar de que la imagen se adquirió ortorrectificada, se procedió a verificarla aplicando la técnica definida por Buiten y van Putten (1997) y Vassilopoulou *et al.* (2002), mediante la utilización de un modelo digital de elevación (DEM) con equidistancia de 30 m, así como de puntos de control, principalmente de la red de senderos y caminos, la imagen fue reproyectada al huso 19 Sur, DATUM World Geodetic System84 (WGS84) (Richard 1993, Chuvieco 1997), obteniendo un error cuadrático medio de 0,17 píxeles,

**Cuadro 1.** Estadísticos descriptivos de las variables de rodal. *N*: número de árboles por hectárea, *G*: área basal, *H<sub>dom</sub>*: altura dominante y *V*: volumen bruto.

Descriptive statistics of plotwise variables per stand. *N* is number of trees per hectare, *G* is stand basal area, *H<sub>dom</sub>* is dominant height, and *V* is stand gross volume.

Estadístico	Rodal 1 ( <i>n</i> = 66)				Rodal 2 ( <i>n</i> = 8)			
	<i>N</i> (árboles/ha)	<i>G</i> (m <sup>2</sup> /ha)	<i>H<sub>dom</sub></i> (m)	<i>V</i> (m <sup>3</sup> /ha)	<i>N</i> (árboles/ha)	<i>G</i> (m <sup>2</sup> /ha)	<i>H<sub>dom</sub></i> (m)	<i>V</i> (m <sup>3</sup> /ha)
Mínimo	180,0	46,5	19,8	291,4	190,0	23,3	11,0	110,3
Máximo	780,0	181,7	39,0	2.078,5	580,0	94,4	22,7	642,9
Media	351,7	100,2	32,6	1.015,1	340,0	43,2	18,2	277,7
Coef. Var. (%)	34,0	27,0	13,0	34,8	40,2	53,6	22,9	62,1



**Figura 1.** Distribución diamétrica por especie y para todas las especies combinadas (recuadro) por rodal.

Diameter distribution for all live trees by species and for all species combined (inset) by stand.

equivalente a 5 m. Aunque la imagen está libre de nubes y la información de los píxeles es de alta calidad, sin ruido, en toda el área de estudio, se utilizó la corrección atmosférica TOAR (*top of atmosphere reflectance*) propuesta por Markham y Chander (2003), la cual ha sido replicada con éxito por Hall *et al.* (2006) sobre las bandas del infrarrojo. Esto corrigió las variaciones en la iluminación solar y asumió la uniformidad de la atmósfera dentro de la imagen. Los valores digitales de la imagen se transformaron a valores de reflectancia, utilizando para ello los valores de calibración de la imagen, información complementaria radiométrica, la distancia Tierra-Sol y ángulo cenital solar; el procedimiento de sustracción de objetos oscuros también se aplicó para reducir los efectos de radiación en las bandas del visible (Teillet y Fedosejevs 1995, Chávez 1996). La corrección topográfica permitió compensar las diferencias de iluminación solar que presentó la imagen, debido a la variación altitudinal del terreno (Conese *et al.*

1993). Con el fin de minimizar este efecto, se realizó la corrección basada en un modelo de iluminación derivado del modelo digital de elevación (Itten y Meyer 1993). Los procesamientos de las imágenes se realizaron utilizando los *software* Idrisi Kilimanjaro (Eastman 2003) y Erdas Imagine versión 8.7 (Leica Geosystems 2005).

*Mediciones de terreno.* Se establecieron 66 y ocho parcelas de muestreo en los rodales 1 y 2, respectivamente. Todas las parcelas fueron de 50 × 20 m (1.000 m<sup>2</sup>). Se identificaron todas las especies de árboles vivos del rodal, con diámetro a la altura del pecho (*d*) mayor a 5 cm, y se midió el *d* y la clase de copa. La altura total (*h*) fue medida en una muestra aleatoria de 7-10 árboles por parcela. Para los árboles no muestra se predijo la altura mediante ecuaciones ajustadas en el presente estudio específicamente para cada especie. El volumen bruto fue estimado con ecuaciones ajustadas para *A. araucana*, *N.*

*dombeyi* y *Nothofagus pumilio* (Poepp et Endl.) Krasser por Siebert (1995), Corvalán (1998) y Salas (2002) respectivamente. Para el resto de las especies se utilizó un factor de forma de 1/3. Los datos calculados por parcela fueron los siguientes: número de árboles por hectárea ( $N$ ), área basal por hectárea ( $G$ ), altura dominante (altura media de los 100 árboles de mayor diámetro en una hectárea) ( $H_{dom}$ ) y el volumen bruto por hectárea ( $V$ ). La estadística descriptiva de las variables objetivo, segregada por rodal, se muestra en el cuadro 1.

Cada parcela fue georreferenciada utilizando un GPS 60CSx Garmin en Datum WGS84 y huso 19 Sur. Las variables altitud, exposición y pendiente, fueron determinadas para cada píxel utilizando un modelo digital de elevación (Brockhaus y Khorram 1992, Ustin y Trabucco 2000). Se calcularon las coordenadas UTM del centro de la parcela para cada parcela de terreno. Los valores correspondientes de las variables de Landsat (banda 1 y así sucesivamente) para cada parcela son una suma de la proporción del valor de cada píxel que es cubierto para el área de la parcela. De esta manera, las variables auxiliares  $X$  contenidas por el píxel y las variables de interés  $Y$  medidas a nivel de parcela son vinculadas espacialmente. El error de la georreferenciación de las parcelas es 4-6 m, lo que adicionado al error de ortorrectificación de la imagen entregó un error de 11 m. Para la extracción de la respuesta espectral se utilizó una ventana de píxeles  $3 \times 3$ , de la cual se extrajo el valor espectral medio por banda, derivado de la ponderación de acuerdo al grado de participación del valor digital en relación a la superficie de la parcela, privilegiando el píxel central, a diferencia de Lu *et al.* (2004) y Hall *et al.* (2006), donde sólo se consideró el valor medio; de esta manera se redujo la diferencia de sensibilidad de la imagen derivada del desfase geométrico.

**Métodos estadísticos.** Se evaluó el desempeño de los métodos antes mencionados para la predicción de las siguientes variables a nivel de parcela:  $N$  (en árboles por hectárea),  $G$  (en  $m^2/ha$ ),  $H_{dom}$  (en m) y  $V$  (en  $m^3/ha$ ). Las variables de rodal de las parcelas son denotadas por  $Y$  y los valores asociados a las bandas derivadas de la imagen Landsat son denotadas por  $X$ . Se utilizaron las bandas originales procesadas de Landsat-7 ETM+, 1-5 y 7. Los métodos fueron ajustados usando una base de datos con observaciones para las cuales se tienen los valores de las variables predictoras y respuestas. Esta base de datos en la terminología de sensores remotos se denomina de “entrenamiento” o de “referencia” y en terminología estadística de “ajuste”. Las observaciones de los datos de referencia son indexados por  $j$ , donde  $j = 1 \dots n_r$  y  $n_r$  es el tamaño muestral de los datos. Se evaluaron los métodos en la predicción de la variable respuesta de interés ( $Y$ ) en una base de datos con observaciones donde sólo los valores de  $X$  son conocidos. Esta base de datos en la terminología de sensores remotos se denomina

de *test*, y en terminología estadística “validación”. Las observaciones de los datos de *test* son indexados por  $u$ , donde  $u = 1 \dots n_t$  y  $n_t$  es el tamaño muestral de los datos. Para los métodos no paramétricos fueron usadas todas las variables  $X$ , mientras que sólo una porción del total de bandas fue empleada en los métodos paramétricos, para evitar multicolinealidad. Es importante también mencionar que aunque los métodos no paramétricos que se evalúan pueden ser utilizados para la predicción de variables cualitativas, el presente estudio se centró en la predicción de variables cuantitativas.

- a) Regresión lineal múltiple (MLR). Se ajustó a un modelo de regresión lineal para cada variable respuesta, utilizando bandas seleccionadas como variables predictoras, y el rodal como una variable muda (*dummy*). La forma general del modelo lineal, en notación matricial, es [1]:

$$y = X\beta + \epsilon, \quad [1]$$

donde  $y$  es un vector con las observaciones de terreno de cada parcela;  $X$  es una matriz que contiene las variables predictoras;  $\beta$  es un vector con los parámetros del modelo a ser estimados y  $\epsilon$  es un vector aleatorio que contiene los errores asociados con cada observación y es distribuido como  $N(0, \sigma^2 I)$  (normal con media 0 y una matriz varianza-covarianza  $\sigma^2 I$ ), donde  $\sigma^2$  es la varianza de los residuales y  $I$  es una matriz de identidad. Se obtuvieron así los parámetros estimados usando mínimos cuadrados ordinarios (*ordinary least squares*, OLS).

Se seleccionaron las variables predictoras al revisar la correlación entre ellas para cada una de las variables respuesta, y por pruebas de significancia estadística de la variable predictora, empleando una simple prueba- $t$  del correspondiente parámetro que fue estimado. Se evitó deliberadamente el uso de la regresión paso a paso, un proceso automático, por ser un método sesgado (Thompson 1995, Grossman *et al.* 1996). Las variables predictoras fueron rodal (como variable muda) y bandas 3 y 7 (denominadas como *banda3* y *banda7*, respectivamente) para todas las variables respuesta, con la excepción del modelo de  $N$  donde *banda7* fue lo suficientemente parsimonioso. Las predicciones se obtuvieron al emplear el vector OLS de parámetros estimados ( $\hat{\beta}_{OLS}$ ) en el modelo lineal [1].

- b) Mínimos cuadrados generalizados (GLS). En un modelo GLS se generaliza el supuesto de distribución de OLS desde  $Cov-Var(\epsilon) = \sigma^2 I$  a  $Cov-Var(\epsilon) = \sigma^2 \Sigma$ , donde  $\Sigma$  es alguna matriz simétrica. Los datos se obtuvieron desde ubicaciones espaciales y en OLS se asumió que los errores de los modelos son independientes. Sin embargo, se espera que los datos que se encuentren más cerca serán más parecidos que los que se encuentren más lejos, lo cual implica correlación espacial o autocorrelación.  $\Sigma$  es la matriz de covarianza que permite tomar en consideración la correlación entre

observaciones al obtener los parámetros. Esta matriz se construye para representar la correlación entre observaciones, dependiendo de la distancia entre los centroides de las parcelas. Las estructuras espaciales de correlación son generalmente representadas por sus funciones de semivariograma en vez de sus funciones de correlación (Waller y Gotway 2004). Se evaluó el uso de los siguientes modelos de semivariogramas, como los descritos en Pinheiro y Bates (2000), exponencial, gaussiano, lineal, cuadrático-racional y esférico. También se evaluó el efecto de emplear un efecto *nugget* en los semivariogramas. Los modelos fueron comparados mediante el valor del criterio de información de Akaike (*Akaike information criterion*) (AIC) (Akaike 1973), el cual se basa en el valor maximizado del logaritmo de la verosimilitud del modelo y en el número de parámetros, del modelo. El ajuste del modelo fue realizado en el *software* estadístico R (R Development Core Team 2008).

c) Modelo lineal de efectos mixtos (LME). Un modelo lineal de efectos mixtos se representa por [2]:

$$y = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \epsilon, \quad [2]$$

donde,  $\boldsymbol{\beta}$  denota un vector de parámetros poblacionales (fijos) desconocidos de dimensión  $p \times 1$ ,  $\mathbf{X}$  una matriz de diseño de dimensión  $n \times p$  relacionando  $\boldsymbol{\beta}$  a  $\mathbf{y}$ ,  $\mathbf{b}$  denota un vector  $k \times 1$  de efectos desconocidos individuales (aleatorios) y  $\mathbf{Z}_i$  una matriz de diseño conocida  $n_i \times k$  relacionando  $\mathbf{b}$  a  $\mathbf{y}$  (Laird y Ware 1982). El vector  $\epsilon$  es independientemente distribuido como  $N(\mathbf{0}, \mathbf{R})$  (normal con media  $\mathbf{0}$  y matriz de covarianza  $\mathbf{R}$ ), y los efectos aleatorios  $\mathbf{b}$  son también distribuidos  $N(\mathbf{0}, \mathbf{D})$  (normal con media  $\mathbf{0}$  y matriz de covarianza de los efectos aleatorios  $\mathbf{D}$ ). Se agregaron efectos aleatorios al intercepto del modelo, es decir, cada parcela tiene una desviación del valor del parámetro de efectos fijos. La matriz de diseño del modelo lineal de efectos mixtos (para cada variable de rodal) contiene las mismas variables predictoras usadas para la regresión lineal múltiple. El modelo lineal de efectos mixtos fue ajustado por máxima verosimilitud (ML), usando el paquete “nlme” (Pinheiro *et al.* 2008) implementado en R. También se ajustó el modelo lineal de efectos mixtos permitiendo correlación espacial en la matriz de covarianza-varianza, tal como fue descrito para mínimos cuadrados generalizados, y luego se evaluó si esto mejoró el ajuste al analizar el AIC como estadístico

Cualquier modelo de efectos mixtos implica múltiples estrategias para predicción. Se puede emplear sólo la parte de los efectos fijos del modelo lineal de efectos mixtos [2] para predicción o usar tanto los efectos fijos y los aleatorios del modelo. Desde aquí en adelante se referirá a la parte fija de [2] como modelo lineal de efectos mixtos y al modelo usando la parte fija y aleatoria de [2] como el modelo lineal

de efectos mixtos con efectos aleatorios conocidos, esto es el nivel 1 de efectos aleatorios o el *innermost* efecto aleatorio. Aunque para los datos usados en el estudio se conocen los efectos aleatorios (éstos son predichos en el ajuste del modelo) y, por lo tanto, estos efectos y los fijos se pueden emplear para predicción, esto no es posible de realizar en la práctica (aplicación del modelo) porque los efectos aleatorios son desconocidos. En este último contexto, se ajustó un modelo complementario que permite predecir estos efectos aleatorios. Se asumió que estos efectos aleatorios estarían relacionados a variables de sitio de la parcela. Se podría usar el modelo de variables de sitio propuesto por Stage y Salas (2007), sin embargo, dado lo restringido de los datos, se prefirió ajustar un modelo más simple empleando las mismas variables como efectos principales. El modelo complementario ajustado fue [3]:

$$b_{0_i} = \beta_0 + Rodal2_i + \beta_1 SL_i + \beta_2 EL_i + \beta_3 ASP_i, \quad [3]$$

donde,  $b_{0_i}$  es el efecto-aleatorio debido a la  $i$ -ésima observación para el intercepto del modelo LME en [2],  $SL$  = pendiente (%),  $EL$  = altitud (m),  $ASP$  = exposición en grados desde el norte, y  $Rodal2$  es una variable muda que toma el valor de 1 cuando la  $i$ -ésima observación pertenece al rodal 2 o 0 en el caso contrario. Stage y Salas (2007) originalmente propusieron el uso de la variable de tipo de hábitat, la cual se ha cambiado acá por la variable muda de rodal, la cual podrá, a grandes rasgos, aproximarse como un subtipo forestal. La variable  $Rodal2$  podrá ser generalizada para representar bosques con baja cobertura y altura, con vegetación en parches, y con una mezcla de especies en el dosel superior. Por otra parte, un bosque similar al rodal 1 ( $Rodal2 = 0$  en [3]) representaría al bosque más común de *A. araucana*, entre los 900 y 1.300 m s.n.m. En la práctica y en ausencia de más datos para poder predecir  $Y$ , un usuario puede emplear sólo los parámetros estimados de los efectos fijos del modelo lineal de efectos mixtos más la predicción de los efectos aleatorios (al usar [3]). Se llamó a esta estrategia para el modelo de efecto mixto como LME-Sys (efectos mixtos lineales con efectos aleatorios mediante un modelo complementario).

d) Mínimos cuadrados parciales (PLS). Mínimos cuadrados parciales es una técnica estadística multivariante usada en disciplinas que emplean una gran cantidad de variables predictoras correlacionadas (Kroonenberg 1990, Wold 2001, Mevik y Cederkvist 2004). Se emplearon mínimos cuadrados parciales para predecir simultáneamente todas las variables  $Y$  en estudio, al crear una matriz  $\mathbf{Y}$  de tamaño  $n \times q$ , donde  $q$  es el número total de variables respuesta, y la matriz de respuestas  $\mathbf{X}$  de tamaño  $n \times p$ ; entonces se ajustó un modelo lineal multivariante como sigue ([4] a [10]):

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{F}, \quad [4]$$

donde  $\mathbf{B}$  es una matriz de coeficientes y  $\mathbf{F}$  es una matriz de errores aleatorios. PLS es similar a componentes principales en el sentido de que produce nuevas variables (o puntajes) que son combinaciones lineales de los predictores originales, como

$$\mathbf{T} = \mathbf{XW}, \quad [5]$$

donde  $\mathbf{W}$  es una matriz  $p \times c$  de ponderaciones de la matriz de predictores  $\mathbf{X}$  de tamaño  $n \times p$ , de tal manera que  $\mathbf{T}$  y  $\mathbf{X}$  son ortogonales, y las columnas de la matriz  $\mathbf{T}$  son llamadas “variables latentes” o “componentes latentes” (Boulesteix y Strimmer 2007). Los mínimos cuadrados parciales se basan en la siguiente descomposición del componente latente:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}, \quad [6]$$

$$\mathbf{Y} = \mathbf{TQ}' + \mathbf{F}, \quad [7]$$

donde  $\mathbf{P}$  y  $\mathbf{Q}$  son, respectivamente,  $p \times c$  y  $q \times c$  matrices de coeficientes (*loading*), y matrices  $\mathbf{E}$  y  $\mathbf{F}$  son los errores. En [7] los componentes latentes son usados para predecir en lugar de las variables originales (Boulesteix y Strimmer 2007) y la matriz de coeficientes  $\mathbf{Q}$  de la ecuación se obtiene al usar OLS, como

$$\hat{\mathbf{Q}} = (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{Y} \quad [8]$$

Finalmente, la matriz  $\mathbf{B}$  para el modelo en [4] está dada por

$$\hat{\mathbf{B}} = \mathbf{W}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{Y} \quad [9]$$

y la matriz respuesta ajustada podrá ser escrita con [7] como:

$$\hat{\mathbf{Y}} = \mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{Y} \quad [10]$$

Según lo destacado por Boulesteix y Strimmer (2007), en PLS la reducción de dimensiones y regresión se llevaron a cabo simultáneamente. Se seleccionaron los primeros cuatro componentes para ambos rodales, después de realizar una validación cruzada (Mevik y Wehrens 2007). El modelo PLS fue ajustado por medio del paquete “pls” de Mevik y Wehrens (2007) implementado en R.

e)  $k$ -ésimo vecino más cercano ( $k$ -NN). El método del  $k$ -ésimo vecino más cercano, también referido como el método de la parcela de muestreo de referencia (Reese *et al.* 2002, Tomppo *et al.* 2008), es un método no paramétrico. El objetivo es predecir los valores de  $Y$  para los datos objetivo basado en alguna similaridad entre las variables  $X$  desde los datos objetivo y de referencia. En otras palabras, la observación objeto  $u$  recibirá un valor  $\hat{y}_u$  derivado desde la  $k$ -ésima observación más

cercana, donde la cercanía entre observaciones de referencia y objetivo se mide en el espacio  $X$  que tiene  $p$  dimensiones, donde  $p$  es el número de variables predictoras. Por lo tanto, una parcela es representada como un vector de dimensión  $1 \times p$  en este espacio, y se denominará vector en un espacio característico multidimensional. La cercanía o variable de distancia en el  $p$  dimensional espacio característico entre el punto objetivo  $u$  y el de referencia  $j$  pueden ser expresados usando la distancia como [11]:

$$d_{uj} = \sqrt{(\mathbf{x}_u - \mathbf{x}_j)(\mathbf{x}_u - \mathbf{x}_j)',} \quad [11]$$

donde  $\mathbf{x}_u$  es el vector caracterizado de dimensión  $1 \times p$  para el  $u$ -ésimo punto de los datos objetivos,  $\mathbf{x}_j$  es el vector caracterizado de los  $j$  puntos de los datos de referencia.  $d_{uj}$  es un escalar que expresa que tan “cerca” el  $j$ -ésimo punto de referencia y el  $u$ -ésimo punto objetivo están en el espacio caracterizado, y todos los  $js$  puntos de referencia pueden ser ordenados en un vector simbolizado por  $\mathbf{d}_{(uj)}$  de dimensión  $n_r \times 1$ .

El procedimiento general para el  $k$ -ésimo vecino más cercano puede ser descrito como sigue: i) calcular la distancia euclideana entre puntos objetivo y de referencia como en [11]; ii) seleccionar un set de los  $k$  puntos de referencia más cercanos mediante algún criterio; iii) ordenar dichas distancias en orden ascendente, simbolizando el vector ordenado por  $\mathbf{d}_{(uj)}$ ; iv) calcular las ponderaciones normalizadas para las primeras  $k$  distancias en el vector  $\mathbf{d}_{(uj)}$ , usando una función de ponderación inversamente proporcional a la raíz cuadrada de la distancia [12]:

$$\mathbf{w}_{(uj)} = \frac{1}{\mathbf{d}_{(uj)}^2} \Big/ \sum_{(uj)=1}^k \left( \frac{1}{\mathbf{d}_{(uj)}^2} \right), \quad [12]$$

y, finalmente, (v) predecir los valores  $y$  para el  $u$ -ésimo punto objetivo con [13]:

$$\hat{y}_u = \sum_{(uj)=1}^k \mathbf{w}_{(uj)} \times y_{(uj)} \quad [13]$$

El método del  $k$ -ésimo vecino más cercano puede ser afinado al encontrar el valor óptimo de  $k$  (número de vecinos),  $k^*$ , y al usar un criterio de distancia apropiado. Dado que en el presente estudio las predicciones se realizan a nivel de parcela, se intentó determinar un  $k^*$  para cada uno de los rodales. Sin embargo, el bajo tamaño muestral en rodal 2 no fue apropiado para determinar  $k^*$ , por lo tanto, sólo el rodal 1 ( $n = 66$ ) fue empleado para estos fines. Se definió  $k^*$  como aquel  $k$  que minimizó la raíz cuadrada de las diferencias entre los valores predichos y observados, usando el método de validación cruzada. Se obtuvo  $k^* = 5$ , valor que también ha sido encontrado en los estudios de

Reese *et al.* (2002) y Kajisa *et al.* (2008); por lo tanto, fue el empleado para ambos rodales. Finalmente, no se encontraron mayores diferencias al usar distintos criterios de distancia que la simple distancia euclideana [11], lo cual coincide con lo reportado por Franco-Lopez *et al.* (2001), Katila y Tomppo (2001), Reese *et al.* (2002) y Kajisa *et al.* (2008), por lo tanto, dicha distancia fue empleada para ambos rodales.

Las variables cualitativas también pueden ser predichas usando las más frecuentes clases con la mayor ponderación total como en Gjertsen (2007). Mayores detalles sobre implementaciones computacionales del método del  $k$ -ésimo vecino más cercano son dados en Finley y McRoberts (2008). Este método ha sido ampliamente estudiado (Ek *et al.* 1997, Reese *et al.* 2002) y ha sido usado operacionalmente en Finlandia desde 1990 (Tomppo 1991). Una revisión detallada del desarrollo del método en el inventario nacional forestal de Finlandia es dado en Tomppo *et al.* (2008).

- f) Vecino más similar (MSN). El método del vecino más similar, propuesto por Moeur y Stage (1995), se basa en un criterio de distancia que es igual a la distancia euclideana ponderada por una matriz  $W$ , como sigue [14, 15],

$$d_{uj} = \sqrt{(\mathbf{x}_u - \mathbf{x}_j) \mathbf{W} (\mathbf{x}_u - \mathbf{x}_j)'}, \quad [14]$$

con :

$$\mathbf{W} = \mathbf{\Gamma} \mathbf{\Lambda}^2 \mathbf{\Gamma} \quad [15]$$

donde  $\mathbf{\Gamma}$  es la matriz de coeficientes canónicos para las observaciones de referencia con dimensión  $p \times s$ , siendo  $s$  el menor valor entre el número de variables  $Y$ :  $q$ , y variables  $X$ :  $p$ ,  $\mathbf{\Lambda}^2$  es la matriz diagonal de correlaciones cuadradas canónicas con dimensiones  $s \times s$ . Packalén y Maltamo (2007) describieron el método  $k$ -MSN, donde  $k$ -vecinos más similares son promediados usando las ponderaciones calculadas en forma inversamente proporcional con la distancia al vecino más similar. Debido a que sólo la primera correlación canónica fue significativa para los datos acá empleados, se restringieron los análisis para el primer componente canónico. Sin embargo, si no hay relación significativa entre las variables  $Y$  e  $X$ , la matriz  $W$  puede ser calculada como la matriz inversa de covarianza de las variables  $X$ , resultando un criterio igual a la distancia Mahalanobis (Crookston *et al.* 2002).

*Comparación.* Se empleó la raíz cuadrada de las diferencias entre los valores predichos y observados para evaluar el comportamiento predictivo de los métodos. Stage y Crookston (2007) y Hudak *et al.* (2008) destacaron que para métodos de vecino más cercano ( $k$ -NN y MSN) el término RMSE (*root mean square errors*) no es apropiado ya que el error del modelo no es resultado de minimizar

la raíz cuadrada de las diferencias entre dos conjuntos de datos. Por el contrario, estos autores recomendaron usar el término RMSD (*root mean square difference*) entre valores predichos y observados. En otros trabajos similares ha sido empleada la raíz cuadrada del error de las predicciones (RMSEP) (Mevik y Wehrens 2007). Se ajustaron y validaron los métodos usando la técnica de validación cruzada *leave-one-out*, que consiste en ajustar los métodos usando  $n-1$  observaciones y probarlos en la observación excluida, y repitiendo este proceso hasta que todas las observaciones hayan sido usadas una vez para probar los métodos. Para evaluar el comportamiento de los métodos bajo estudio se calcularon los estadísticos diferencia media (MD) y RMSD, como sigue [16, 17]:

$$MD_j = \frac{1}{n} \sum_i^n (y_i - \hat{y}_{ij}) \quad [16]$$

donde  $MD_j$  es la diferencia media para el  $j$ -ésimo método evaluado,  $y_i$  es el valor observado de la variable  $y$  en la  $i$ -ésima parcela,  $\hat{y}_{ij}$  es el valor predicho de la variable  $y$  en la  $i$ -ésima parcela para el  $j$ -ésimo método, y

$$RMSD_j = \sqrt{\frac{1}{n} \sum_i^n (y_i - \hat{y}_{ij})^2} \quad [17]$$

Los estadísticos de predicción (MD y RMSD) fueron calculados en términos porcentuales al dividirlos por el valor observado promedio.

## RESULTADOS

Los rodales 1 y 2 tuvieron diferente composición y estructura. Todas las variables de rodal presentaron una mayor variación en el rodal 2 que en el rodal 1 (cuadro 1). Aunque el muestreo fue con distintos tamaños de muestra, se considera que la mayoría de las diferencias entre los rodales se debió a las condiciones de crecimiento. El rodal 1 se ubica en un sitio más productivo y, por lo tanto, tenía un dosel más continuo, mientras que el rodal 2 presentó una discontinuidad y una distribución espacial de copas altamente irregular. Se seleccionaron las bandas 3 y 7 de Landsat como variables predictoras en los modelos de regresión lineal múltiple propuestos para  $G$ ,  $H_{dom}$  y  $V$  (cuadro 2).  $N$  fue la variable más difícil para modelar, debido a la baja correlación lineal con las bandas Landsat. Sin embargo, un modelo que simplemente empleó la banda 7 como variable predictora fue lo suficientemente parsimonioso. Las pruebas de razones de verosimilitud entre regresión lineal múltiple y mínimos cuadrados generalizados y entre modelos lineales de efectos mixtos, usando y no usando una estructura de correlación, mostraron una mejora estadísticamente significativa ( $P < 0,01$ ) para aquellos modelos



**Cuadro 2.** Parámetros estimados de los modelos de regresión lineal múltiple (MLR), mínimos cuadrados generalizados (GLS) y modelo lineal de efectos mixtos (LME), para cada variable de rodal.  $N$  es el número de árboles por hectárea,  $G$  es el área basal del rodal,  $H_{dom}$  es altura dominante y  $V$  es volumen bruto.

Estimated parameters of the multiple linear regression (MLR), generalized least squares (GLS), and linear mixed-effects (LME) models for each stand variable.  $N$  is number of trees per hectare,  $G$  is stand basal area,  $H_{dom}$  is dominant height, and  $V$  is stand gross volume.

Parámetro	MLR				LME				GLS			
	$N$	$G$	$H_{dom}$	$V$	$N$	$G$	$H_{dom}$	$V$	$N$	$G$	$H_{dom}$	$V$
Intercepto	246,881	153,695	43,626	1.823,686	235,513	160,024	30,015	1.523,072	231,477	153,695	30,015	1.523,080
Rodal2	-4,014	-53,874	-13,955	-699,715	3,760	-53,924	-3,979	-597,145	-9,295	-53,874	-3,979	-597,145
Banda3	-	-3,008	-0,519	-41,725	-	-3,275	-0,002	-28,742	-	-3,008	-0,002	-28,742
Banda7	4,090	2,034	0,280	25,636	4,420	2,161	0,012	17,926	4,592	2,034	0,012	17,926

-: sin parametro.

que toman en consideración la correlación espacial. Por lo tanto, el modelo lineal de efectos mixtos presentado aquí tomó en consideración la correlación espacial; sin embargo, según los valores del criterio de información de Akaike, los mínimos cuadrados generalizados tuvieron un mejor ajuste siendo también un modelo estadístico más simple de ajustar. Las estimaciones de parámetros del modelo complementario de efectos aleatorios de cada modelo lineal de efectos mixtos se observan en el cuadro 3.

**Cuadro 3.** Parámetros estimados del modelo complementario [3] que predice los efectos aleatorios de diferentes variables de rodal para el modelo de efectos mixtos.  $N$  es el número de árboles por hectárea,  $G$  es el área basal del rodal,  $H_{dom}$  es altura dominante y  $V$  es volumen bruto. Las variables predictoras son:  $EL$  = altitud (m s.n.m),  $SL$  = pendiente (%) y  $ASP$  = exposición en grados desde el norte.

Site-variables model parameter estimates for the random-effects of the different stand variables for mixed-effects models.  $N$  is number of trees per hectare,  $G$  is stand basal area,  $H_{dom}$  is dominant height, and  $V$  is stand gross volume. Predictor variables are:  $EL$  = elevation (m),  $SL$  = slope (%), and  $ASP$  = aspect in degrees clockwise from north.

Parámetro	Variable de rodal			
	$N$	$G$	$H_{dom}$	$V$
Intercepto	13,054	30,539	-7,093	-12,364
$EL$	-0,019	-0,020	0,007	0,080
$SL$	-0,424	-0,721	-0,053	-7,698
$ASP$	0,102	0,002	-0,008	-0,148

La mayoría de los parámetros estimados entre modelos de regresión lineal múltiple, mínimos cuadrados generalizados y modelo lineal de efectos mixtos (sólo los parámetros de efectos fijos) no difirieron mucho (cuadro 2), sin embargo, la inferencia estadística fue afectada cuando la autocorrelación espacial no fue considerada. Por ejemplo, para el modelo de  $V$ , el ajuste de modelos de regresión

lineal múltiple indicó que todos los parámetros fueron estadísticamente significativos,  $P < 0,05$  (cuadro 4). Por otra parte, con mínimos cuadrados generalizados se obtuvieron parámetros no significativos para las bandas 3 y 7, lo cual haría cambiar drásticamente algún análisis respecto a las variables involucradas.

**Cuadro 4.** Parámetros estimados, errores estándar y valores  $P$  para los modelos de  $V$  mediante regresión lineal múltiple (MLR) y mínimos cuadrados generalizados (GLS).

Estimated parameters, standard errors, and  $P$ -values for both multiple linear regression (MLR) and generalized least squares (GLS) models of  $V$ .

Parámetro	MLR			GLS		
	Valor	Error estándar	$P$	Valor	Error estándar	$P$
Intercepto	1.823,686	450,392	0,000	1.523,080	447,454	0,001
Rodal2	-699,715	125,330	0,000	-597,146	155,989	0,000
Banda3	-41,725	18,383	0,026	-28,742	17,654	0,108
Banda7	25,636	11,420	0,028	17,926	11,385	0,120

Existieron diferencias en el desempeño entre las variables predichas, rodales y métodos estadísticos. Independiente del método que se evaluó, las variables más difíciles de modelar fueron  $N$  y  $V$ , seguidas de  $G$ , y  $H_{dom}$ , respectivamente (cuadro 5). Los errores de predicción fueron mayores en el rodal 2 que en el rodal 1, para todos los métodos y todas las variables. El modelo lineal de efectos mixtos con efectos aleatorios conocidos (LME1) obtuvo el mejor desempeño, con RMSD inferiores a 5% para todas las variables. El desempeño de regresión lineal múltiple fue el segundo mejor para todas las variables en el rodal 1, seguido por mínimos cuadrados parciales,  $k$ -ésimo vecino más cercano, LME-Sys y vecino más similar, respectivamente. Aunque la regresión lineal múltiple obtuvo los mejores resultados

**Cuadro 5.** Estadísticos de predicción para los diferentes métodos estadísticos por rodal. MLR es regresión lineal múltiple, GLS es mínimos cuadrados generalizados, LME1 es modelo lineal de efectos mixtos con efectos aleatorios conocidos, LME-Sys es efectos mixtos lineales con efectos aleatorios predichos, PLS es mínimos cuadrados parciales, *k*-NN es *k*-ésimo vecino más cercano y MSN es vecino más similar.

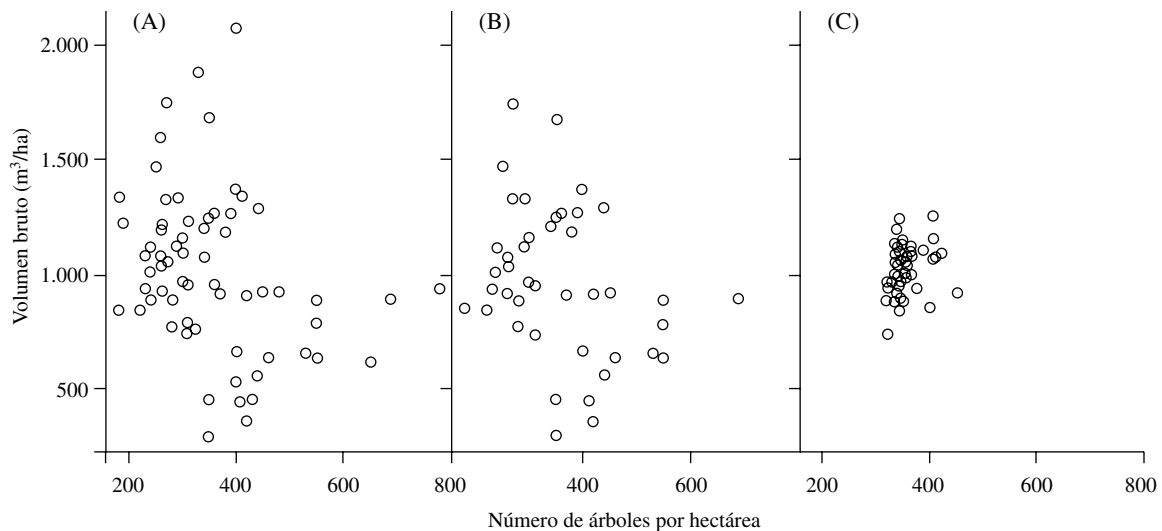
Prediction statistics for different statistical methods for each stand. MLR is multiple linear regression, GLS is generalized least squares, LME1 is linear mixed-effects with known random-effects, LME-Sys is linear mixed-effects with predicted random-effects, PLS is partial least squares, *k*-NN is *k*-nearest neighbor, and MSN is most similar neighbor.

Método	Variable	Rodal 1				Rodal 2			
		MD	MD (%)	RMSD	RMSD (%)	MD	MD (%)	RMSD	RMSD (%)
MLR	<i>N</i>	-0,15	-0,04	119,26	33,91	-0,01	0,00	145,17	42,70
	<i>G</i>	-0,05	-0,05	26,81	26,7	0,10	0,23	25,39	58,74
	<i>H<sub>dom</sub></i>	0,00	0,00	4,23	13,00	0,01	0,07	4,54	24,94
	<i>V</i>	-0,34	-0,03	350,46	34,53	1,03	0,37	195,44	70,39
GLS	<i>N</i>	-0,11	-0,03	116,30	33,07	-25,01	-7,36	125,45	36,90
	<i>G</i>	-0,05	-0,05	25,75	25,71	0,00	0,00	22,05	51,02
	<i>H<sub>dom</sub></i>	0,00	-0,01	4,05	12,43	0,00	0,00	3,95	21,70
	<i>V</i>	-0,44	-0,04	335,48	33,05	0,00	0,00	169,46	61,03
LME1	<i>N</i>	1,82	0,52	64,39	18,31	4,68	1,38	68,56	20,17
	<i>G</i>	0,00	0,00	3,18	3,17	0,00	0,00	2,72	6,29
	<i>H<sub>dom</sub></i>	0,00	0,00	0,50	1,53	0,00	0,00	0,49	2,68
	<i>V</i>	0,00	0,00	41,43	4,08	0,00	0,00	20,87	7,52
LME-Sys	<i>N</i>	-0,03	-0,01	116,27	33,06	0,00	0,00	126,95	37,34
	<i>G</i>	0,00	0,00	25,79	25,75	0,00	0,00	22,04	51,00
	<i>H<sub>dom</sub></i>	0,00	0,00	4,05	12,43	0,00	0,00	3,95	21,70
	<i>V</i>	0,11	0,01	336,11	33,11	-0,41	-0,15	169,35	60,99
PLS	<i>N</i>	-0,56	-0,16	122,34	34,79	13,78	4,05	271,81	79,94
	<i>G</i>	0,19	0,19	27,88	27,84	-3,40	-7,88	40,36	93,38
	<i>H<sub>dom</sub></i>	0,03	0,08	4,49	13,78	-1,93	-10,60	6,51	35,71
	<i>V</i>	2,99	0,29	364,59	35,92	-34,39	-12,39	279,80	100,77
<i>k</i> -NN	<i>N</i>	0,22	0,06	129,25	36,75	20,07	5,90	114,90	33,79
	<i>G</i>	-0,25	-0,25	28,20	28,16	4,72	10,93	28,17	65,18
	<i>H<sub>dom</sub></i>	-0,48	-1,46	4,53	13,90	0,94	5,16	5,05	27,72
	<i>V</i>	-14,70	-1,45	380,20	37,45	33,42	12,04	213,33	76,83
MSN	<i>N</i>	-1,82	-0,52	170,74	48,55	-22,50	-6,62	115,76	34,05
	<i>G</i>	2,49	2,49	36,48	36,42	3,05	7,06	35,77	82,77
	<i>H<sub>dom</sub></i>	0,20	0,61	5,91	18,16	1,61	8,85	5,87	32,20
	<i>V</i>	28,83	2,84	474,32	46,73	29,76	10,72	270,56	97,44

para tres de las cuatro variables en el rodal 2 (excepto para *N*), ni la misma tendencia, ni tendencias claras se pudieron obtener entre los métodos para este rodal. La RMSD para todos los métodos en el rodal 1, excepto para el vecino más similar, fue alrededor de 36 %, 27 %, 14 % y 36 %, para *N*, *G*, *H<sub>dom</sub>*, *V*, respectivamente.

La relación entre las variables originales se mantuvo utilizando el método del vecino más similar. Aunque con regresión lineal múltiple se realizaron buenas predicciones,

la relación entre las mismas variables que se predicen por regresión lineal múltiple fue drásticamente cambiada en comparación con las variables originales. La relación original entre el volumen de rodal y la densidad del rodal (figura 2A) fue casi idéntica a la predicha para las mismas variables utilizando el método del vecino más similar (figura 2B). Por otra parte, la misma relación fue altamente distorsionada cuando las variables fueron predichas utilizando regresión lineal múltiple (figura 2C).



**Figura 2.** Gráficos de dispersión entre volumen bruto y número de árboles por hectárea para los datos observados (A), predicciones del método del vecino más similares (MSN) (B) y predicciones con regresión lineal múltiple (MLR) (C).

Scatterplot between stand gross volume and number of trees per hectare for (A) observed data, (B) most similar neighbor (MSN) predictions, and (C) multiple linear regression (MLR) predictions.

## DISCUSIÓN

La mayoría de los estudios donde se predicen variables de rodal empleando Landsat utilizan datos sobre amplias regiones (5.000-3.000.000 ha) con varias especies (3-25) (Mäkelä y Pekkarinen 2004, Hall *et al.* 2006, Sivanpillai *et al.* 2006, Gjertsen 2007, Lencinas y Mohr-Bell 2007, McRoberts y Tomppo 2007, Kajisa *et al.* 2008). Existe una mayor variabilidad espectral en estos estudios de áreas grandes y la falta de información fiable puede ser un problema, especialmente para países o regiones que no poseen datos de terreno. Por ello, se ha preferido centrar el estudio en un área relativamente pequeña, utilizando rodales representativos de los bosques de *A. araucana* donde pocas especies comparten el dosel superior. Los rangos de topografía y altitud de los rodales son también representativos de la región de Cordillera de Los Andes. Los rodales del presente estudio pueden ser clasificados como homogéneos, siguiendo la definición dada por Mäkelä y Pekkarinen (2004), porque ellos no han sufrido disturbios antropogénicos ni se encuentran en zonas ecotonales. En la presente investigación también se ha reducido la mezcla de información espectral, utilizando parcelas de terreno más grandes (1.000 m<sup>2</sup>) que las comúnmente utilizadas en estudios similares, donde el área de la parcela de terreno es más pequeña que el área cubierta por un píxel.

La estructura y composición de los bosques afectan el desempeño de los métodos estadísticos evaluados en el presente estudio. El rodal 1 presenta una cobertura espacial continua. Por el contrario, el rodal 2 tiene menos cobertura de copas, varios parches de vegetación y una altura de dosel más baja que el rodal 1. En el rodal 1, *N. dombeyi*

es la especie principal del dosel superior, mientras que en el rodal 2 *A. araucana* y *N. antarctica* comparten el dosel superior. La mezcla de especies en el dosel superior del rodal 2 y una cobertura de copas muy inferior contribuyen a una mayor variabilidad espectral en las bandas de Landsat. Esto, además de una mayor variabilidad de las variables medidas en terreno, producirá inevitablemente peores resultados para los bosques de *A. araucana* representados por el rodal 2, en comparación con los bosques representados por el rodal 1. Los resultados del presente estudio están de acuerdo con Poso *et al.* (1984), donde los buenos resultados que obtuvieron pueden estar relacionados con la prevalencia de un dosel cerrado, lo que reduce el efecto de sombra sobre la copa de los árboles dominantes. Además, es difícil estimar las características de los árboles en los estratos no dominantes utilizando técnicas de sensores remotos (Packalén y Maltamo 2007). Finalmente, la estructura y composición del bosque tiene un rol clave en el comportamiento predictivo de los métodos que utilizan variables derivadas de Landsat.

Como se esperaba, la extrapolación de los resultados para el rodal 2 resulta problemática. Los resultados son muy claros para el rodal 1, el cual tiene una superficie mayor y un mayor número de parcelas. Se observó un desempeño consistentemente mejor de los métodos entre las variables del rodal 1. Esta tendencia no se mantiene para el rodal 2, donde no se puede seleccionar con claridad entre los métodos estadísticos. El reducido tamaño muestral utilizado para el rodal 2 y su estructura y composición han contribuido a un rendimiento más pobre de los métodos en este rodal. Este resultado era esperado, porque el rodal 2 es un subtipo de bosque con una estructura más cercana

a arbustos que a un bosque de dosel cerrado, donde las técnicas de sensores remotos no ofrecen información de manera confiable para los objetivos planteados en este estudio. No obstante, estos resultados parecen ser realistas para este tipo de comunidad vegetal. Aunque un mayor tamaño de la muestra puede reducir los errores, no se espera una reducción significativa de errores de predicción.

Los métodos estadísticos aquí evaluados deben ser aplicados con precaución en áreas similares con otras imágenes Landsat. Tanto los métodos paramétricos como los no paramétricos pueden ser utilizados para predecir  $Y$  para píxeles donde esa variable es faltante, pero solo en la misma imagen Landsat, por ejemplo, para rodales cubiertos por la imagen acá empleada, pero no fueron utilizados para el ajuste de los métodos. Para utilizar estos modelos ya ajustados para un tipo de bosque similar empleando nuevos datos (diferente imagen Landsat), la imagen debe mostrar propiedades espectrales similares a la utilizada en la parametrización del modelo (fecha similar de colección/estación y similares condiciones atmosféricas/corrección). Si éstos y otros requisitos se cumplen, los únicos métodos que serán adecuados son los paramétricos.

El modelo lineal de efectos mixtos con efectos aleatorios conocidos (LME1) exhibió el mejor comportamiento predictivo para todas las variables. En la práctica, esto es, aplicar el modelo ajustado sin poseer los datos empleados para su ajuste, lo más probable es que estos resultados no serán alcanzables. Mientras el RMSD obtenido por el modelo lineal de efectos mixtos con efectos aleatorios conocidos es menor que el resto de los métodos, el modelo ajustado por sí mismo no puede ser aplicado a otros datos, porque los efectos aleatorios se desconocen. Los problemas surgen cuando se hacen las predicciones para distintas áreas de estudio y se requieren datos, ya que tanto el modelo lineal de efectos mixtos con efectos aleatorios conocidos, el  $k$ -ésimo vecino más cercano y el vecino más similar no pueden ser usados. Aunque sólo los métodos paramétricos (con la excepción de LME1) pueden ser utilizados para rodales similares cubiertos por imágenes Landsat diferentes, tal como se explicó en el párrafo anterior, parece que en estudios de sensores remotos la atención se centra en realizar predicciones para la imagen que se está analizando. En este contexto, el modelo lineal de efectos mixtos con efectos aleatorios conocidos sería la mejor alternativa. Una conclusión similar encontraron Salas *et al.* (2010) empleando variables derivadas con LiDAR para modelar diámetros de árboles individuales en Noruega. Además, si se dispone de algunos datos (que no fueran suficientes para el ajuste de un nuevo modelo) para una nueva área de estudio, los modelos de efectos mixtos ofrecen una manera de predecir los efectos aleatorios (o calibrar) a través de los BLUPs (best linear unbiased predictors) (Robinson 1991). El modelo de efectos mixtos lineales con efectos aleatorios predichos (LME-Sys) surge como una alternativa al modelo lineal de efectos mixtos con efectos aleatorios conocidos, ya que los efectos aleatorios del modelo original de efectos mixtos son predichos

mediante un modelo auxiliar [3] (parámetros estimados aparecen en el cuadro 3).

Luego del modelo lineal de efectos mixtos con efectos aleatorios conocidos, los mínimos cuadrados generalizados exhibieron el mejor comportamiento predictivo para tres de las cuatro variables respuesta analizadas (para  $N$ , LME-Sys fue mejor). Los mínimos cuadrados generalizados son un modelo más simple de ajustar que el de efectos mixtos lineales con efectos aleatorios predichos (LME-Sys) y directamente toma en consideración la correlación espacial, por lo tanto, se prefiere. La regresión lineal múltiple es el modelo paramétrico más simple evaluado en el presente estudio y tuvo una buena capacidad predictiva, con rangos de error similares a los reportados por Eckert *et al.* (2005) en bosques naturales de *N. pumilio* en Argentina, para área basal y volumen, pero con imágenes ASTER. No obstante, los supuestos estadísticos de regresión lineal múltiple no se cumplen. Los mínimos cuadrados parciales se comportan muy similares, seguidos por el  $k$ -ésimo vecino más cercano. Sólo se encontraron claras diferencias con el vecino más similar. Se esperará un mejor comportamiento del vecino más similar sobre áreas más extensas.

Si sólo se consideran los resultados obtenidos para el rodal 1, existen diferencias en el desempeño predictivo entre los métodos paramétricos y no paramétricos. El desempeño de los métodos paramétricos fue muy bueno y sus resultados no difirieron significativamente entre sí. Tanto los mínimos cuadrados generalizados y el modelo de efectos mixtos lineales con efectos aleatorios predichos alcanzaron bajos (no considerando LME1) RMSD y MD para todas las variables. Las diferencias para RMSD entre mínimos cuadrados generalizados y mínimos cuadrados parciales son de alrededor de 1-2%, y exhiben un comportamiento muy similar. Aunque la regresión lineal múltiple es el modelo más simple, sólo es univariante y necesita ser ajustado para cada variable respuesta. Por otra parte, los mínimos cuadrados parciales son un método multivariado y necesita ser ejecutado una sola vez para una base de datos dada. A diferencia de los métodos paramétricos, el desempeño de los dos métodos no paramétricos fue significativamente diferente entre sí. Las diferencias en RMSD entre  $k$ -ésimo vecino más cercano y vecino más similar fueron de alrededor del 10% para el rodal 1 en todas las variables. Mediante el  $k$ -ésimo vecino más cercano, se obtuvieron errores en el volumen de rodal (37,45%) bastante más bajos que el 66% reportado por Kajisa *et al.* (2008). La diferencia podría estar dada por la mayor área cubierta por Kajisa *et al.* (2008), así como su mayor variabilidad (34,8% versus 77,3% para  $V$  en Kajisa *et al.* (2008)).

Los métodos paramétricos y no paramétricos ofrecen una solución interesante para diferentes preguntas de investigación. Los métodos paramétricos ofrecen un marco apropiado para pruebas de hipótesis estadísticas, pero requieren el cumplimiento de varios supuestos. Uno de los problemas del  $k$ -ésimo vecino más cercano y vecino más similar es respecto del uso de los métodos en otras áreas que las usadas para ajustar los métodos. Mayores

detalles sobre estadística paramétrica y no paramétrica se pueden revisar en diversos trabajos (Potvin y Roff 1993, Baíllo y Cuevas 2003, Bissantz *et al.* 2003, Vickers 2005, Weinand 2007), pero usualmente en ecología se tiene a inclinarse por las ventajas de los métodos no paramétricos (véanse las revisiones de Johnson 1995, Smith 1995, Stewart-Oaten 1995). El modelo lineal de efectos mixtos con efectos aleatorios conocidos (LME1) es la mejor opción para los datos empleados en el presente estudio (rodal 1), sin embargo, no puede ser utilizado para un área de estudio diferente donde no existen datos disponibles. Aunque los modelos de regresión lineal múltiple parecen ser la mejor opción para la predicción de las variables de rodal desde variables derivadas de Landsat, sus supuestos no se cumplen para datos de sensores remotos que son correlacionados espacialmente. Esto afecta la inferencia estadística de los modelos ajustados por regresión lineal múltiple como, por ejemplo, la prueba de hipótesis para la significancia estadística de algunos parámetros no sería correcta. El mismo problema ocurre para los mínimos cuadrados parciales, aunque estos ofrecen el ajuste simultáneo de los modelos, siendo sus predicciones consistentes (o, a veces, llamados compatibles) entre los modelos, lo cual no es posible cuando la regresión lineal múltiple es ajustada para cada variable respuesta por separado. Las predicciones del vecino más similar mantienen las relaciones originales entre las variables, lo cual es importante cuando es necesario hacer predicciones para otros sectores o extrapolaciones. Además, los métodos no paramétricos tienen la ventaja de que simultáneamente predicen variables cualitativas y cuantitativas (como se mencionó anteriormente). Como una regla general, se podría pensar en modelos paramétricos que son especialmente adecuados para comparaciones estadísticas; por otra parte, los métodos no paramétricos parecieran tener un mejor desempeño para propósitos de predicción. Sin embargo, un modelo paramétrico (LME-Sys) cumple con los supuestos estadísticos y alcanza niveles razonables de errores. Además, existen dos ventajas fundamentales que favorecen a los métodos paramétricos: a) ellos entregan estimaciones insesgadas, mientras que los resultados de los métodos no paramétricos dependen del ajuste de la base de datos, la cual debe ser lo suficientemente grande como para cubrir la variabilidad de terreno de las variables. Sin embargo, el tamaño de la base de datos no garantiza insesgamiento. b) Dado lo anterior, ellos podrán usualmente requerir un menor conjunto de datos para el ajuste de los modelos, por lo tanto, ser más económicos.

## CONCLUSIONES

Dada la variabilidad natural de los bosques de *A. araucana* se considera que los resultados son satisfactorios. Se concluye que para el rodal 1 el modelo lineal de efectos mixtos con efectos aleatorios conocidos tiene el mejor comportamiento predictivo. La regresión lineal múltiple, el  $k$ -ésimo vecino más cercano, los mínimos cuadrados

parciales y modelo de efectos mixtos lineales con efectos aleatorios predichos (LME-Sys) se comportan de forma similar entre ellos, pero estos métodos tienen RMSD mucho mayores que el modelo lineal de efectos mixtos con efectos aleatorios conocidos (LME1). Sin embargo, no se puede obtener este tipo de conclusiones desde el rodal 2, debido al pequeño tamaño de la muestra y la mayor variabilidad de la vegetación y las condiciones del sitio. Aunque es importante destacar que tanto los métodos estadísticos paramétricos como los no paramétricos son útiles y su uso depende del objetivo del estudio, a partir de los resultados obtenidos se prefieren los métodos paramétricos. Sin embargo, si el mantenimiento de las relaciones entre las variables originales es una característica atractiva para un investigador, el método del vecino más similar podría ser el apropiado. Por otra parte, la regresión lineal múltiple es una buena opción para los propósitos de predicción, ya que es el más simple y fácil de implementar, pero no se recomienda ya que los requerimientos estadísticos para la aplicación de la regresión lineal múltiple no se cumplen. El modelo lineal de efectos mixtos con efectos aleatorios conocidos (LME1) ofrece la mejor predicción, pero sus excelentes resultados no podrán ser alcanzados, por los mismos modelos acá ajustados, en una nueva imagen de la cual no se dispongan datos de terreno con los cuales calibrar la predicción de los efectos aleatorios. Por otra parte, y aunque el modelo de efectos mixtos lineales con efectos aleatorios predichos (LME-Sys) no logra los mejores resultados, fue lo suficientemente similar que los mejores métodos y representa la estructura de correlación de los datos, por lo tanto, se prefiere.

## AGRADECIMIENTOS

Esta investigación fue financiada por la Dirección de Investigación y Desarrollo de la Universidad de La Frontera, Chile, Proyecto DIUFRO-reg DI07-0057. Damos las gracias a la Corporación Nacional Forestal, especialmente a los guardaparques del Parque Nacional Conguillío, Rodrigo Molina, Fabián Luna, Juan Arriagada y Rodrigo Marín, en la recolección de datos. Por último, agradecemos a Dylan Craven y Zachary Parisa que proporcionaron una revisión crítica de una versión previa de este manuscrito.

## REFERENCIAS

- Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In Petrov BN, G Czaki eds. Second International Symposium on Information Theory. Budapest, Hungary. Akademiai Kiadó. p. 267-281.
- Bahamóndez C, M Martín, S Müller-Using, A Pugin, Y Rojas, G Vergara, O Peña, M Uribe, R Ipinza. 2007. Inventario de los bosques de alerce. Valdivia, Chile. Informe final de Proyecto. Instituto Forestal (INFOR). 166 p.

- Baíllo A, A Cuevas. 2003. Parametric versus nonparametric tolerance regions in detection problems. Working Paper 03-70, Statistics and Econometrics Series 17. Madrid, España. Universidad Carlos III de Madrid, Departamento de Estadística y Econometría. Madrid, Spain.
- Bauer ME, TE Burk, AR Ek, PR Coppin, SD Lime, TA Walsh, DK Walters, W Befort, DF Heinzen. 1994. Satellite inventory of Minnesota forest resources. *Photogramm. Eng. Rem. Sens.* 60: 287-298.
- Bekessy S, AC Newton, JC Fox, A Lara, A Premoli, M Cortés, M González, B Burns, L Gallo, F Izquierdo, M Burgman. 2004. The monkey puzzle tree (*Araucaria araucana*) in southern Chile: effects of timber and seed harvest, volcanic activity, and fire. In Akcakaya R, M Burgman, O Kindvall, C Word, P Sjogren-Gulven, JS Hatfield, MA McCarthy eds. Species Conservation and Management. New York, USA. Oxford University Press. p. 48-63.
- Benoit I. 1996. Representatividad ecológica del sistema nacional de áreas protegidas del Estado. In Muñoz M, H Núñez, J Yáñez eds. Libro rojo de los Sitios Prioritarios para la Conservación de la Diversidad Biológica. Santiago, Chile. Corporación Nacional Forestal. p. 69-94.
- Bissantz N, A Munk, A Scholz. 2003. Parametric versus non-parametric modelling? Statistical evidence based on p-value curves. *Monthly Notice of the Royal Astronomical Society* 340(4):1190-1198.
- Boulesteix AL, K Strimmer. 2007. Partial least squares: a versatile tool for the analysis of high dimensional genomic data. *Briefings in Bioinformatics* 8(1): 32-44.
- Brockhaus JA, S Khorram. 1992. A comparison of SPOT and landsat TM data for use in conducting inventories of forest resources. *Int. J. Rem. Sensing* 13: 3035-3043.
- Buiten HJ, B van Putten. 1997. Quality assessment of remote sensing image registration and testing of control point residuals analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 52: 57-73.
- Burns B. 1993. Fire-induced dynamics of *Araucaria araucana*-*Nothofagus antarctica* forest in the southern Andes. *Journal of Biogeography* 20(6): 669-685.
- Chavez P. 1996. Image-based atmospheric corrections revisited and improved. *Photogramm. Eng. Rem. Sens.* 62: 1025-1036.
- Chuvieco E. 1997. Fundamentos de Teledetección Espacial. Madrid, España. Rialp. 568 p.
- CONAF (Corporación Nacional Forestal, CL), CONAMA (Comisión Nacional del Medio Ambiente, CL), BIRF (Banco Interamericano de Reconstrucción y Fomento). 1999. Catastro y evaluación de recursos vegetacionales nativos de Chile. Proyecto. Santiago, Chile. 86 p. (Informe Novena Región).
- Conese C, MA Gilabert, F Maselli y L Bottal. 1993. Topographic normalization of TM scenes through use of an atmospheric correction method and digital terrain models. *Photogramm. Eng. Rem. Sens.* 59: 1745-1753.
- Corvalán P. 1998. Modelos dendrométricos para la especie *Araucaria araucana* (Mol.). C. Koch en rodales fuertemente intervenidos. *Ciencias Forestales* 12(1-2): 33-41.
- Crookston NL, M Moeur, D Renner. 2002. Users guide to the most similar neighbor imputation program version 2. USA. USDA For. Serv. Gen. Tech. Rep. RMRS-96. 35 p.
- Donoso C. 1981. Tipos forestales de los bosques nativos de Chile. Investigación y Desarrollo Forestal (CONAF/PNUD/FAO), Documento de Trabajo N° 38 (Publicación FAO). 82 p.
- Drake F, MA Herrera, E Acuña. 2005. Propuesta de manejo sustentable de *Araucaria araucana* (Mol. C. Koch). *Bosque* 26(1): 23-32.
- Eastman JR. 2003. IDRISI. user's guide. Worcester, MA, USA. Clark University. 130 p.
- Eckert S, T Kellenberger, JD Lencinas. 2005. Classification and forest parameter extraction of Patagonian Lenga forests with ASTER and Landsat ETM+ Data. In Proceedings PECORA 16. Global Priorities in Land Remote Sensing. Sioux Falls, South Dakota, USA. October 23-27, 2005. 11 p.
- Ek AR, AP Robinson, PJ Radtke, DK Walters. 1997. Development and testing of regeneration imputation models for forests in Minnesota. *For. Ecol. Manage.* 94(1-3): 129-140.
- Finley AO, RE McRoberts. 2008. Efficient k-nearest neighbor searches for multi-source forest attribute mapping. *Remote Sensing of Environment* 112: 2203-2211.
- Franco-Lopez H, AR Ek, ME Bauer. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sensing of Environment* 77: 251-274.
- Gjertsen AK. 2007. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sensing of Environment* 110: 420-430.
- González ME. 2005. Fire history data as reference information in ecological restoration. *Dendrochronologia* 22: 149-154.
- González ME, TT Veblen, J Sibold. 2005. Fire history of *Araucaria-Nothofagus* forests in Villarrica National Park, Chile. *Journal of Biogeography* 32: 1187-1202.
- González M, M Cortés, F Izquierdo, L Gallo, C Echeverría, S Bekkesy, P Montaldo. 2006. *Araucaria araucana*. In Donoso C ed. Las especies arbóreas de los bosques templados de Chile y Argentina. Autoecología. Valdivia, Chile. Marisa Cuneo Ediciones. p. 36-53.
- Grossman YL, SL Ustin, S Jacquemoud, EW Sanderson, G Schmuck, J Verdebout. 1996. Critique of stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. *Remote Sensing of Environment* 56: 182-193.
- Hall RJ, RS Skakum, EJ Arsenault, BS Case. 2006. Estimating stem volume and basal area in forest compartments by combining satellite image data with field data. *For. Ecol. Manage.* 225: 378-390.
- Hudak A, N Crookston, J Evans, D Hall, M Falkowski. 2008. Nearest neighbor imputation of specie-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment* 112: 2232-2245 (and corrigendum in 113: 289-290).
- Itten KI, P Meyer. 1993. Geometric and radiometric correction of TM data of mountainous forested areas. *IEEE Transactions on geoscience and remote sensing* 31: 764-770.
- Johnson DH. 1995. Statistical sirens: The allure of nonparametrics. *Ecology* 76(6): 1998-2000.
- Kajisa T, T Murakami, N Mizoue, F Kitahara, S Yoshida. 2008. Estimation of stand volumes using the k-nearest neighbors method in Kyushu, Japan. *Journal of Forest Research* 13: 249-254.
- Katila M, E Tomppo. 2001. Selecting estimation parameters for the finnish multi-source national forest inventory. *Remote Sensing of Environment* 76(1): 16-32.

- Kroonenberg PM. 1990. Review: Latent variable path modeling with partial least squares. *Journal of the American Statistical Association* 85(411): 909-910.
- Laird NM, JH Ware. 1982. Random-effects models for longitudinal data. *Biometrics* 38(4): 963-974.
- Lee B, SY Kim, J Chung, PS Park. 2008. Estimation of fire severity by use of Landsat TM images and its relevance to vegetation and topography in the 2000 Samcheok forest fire. *Journal of Forest Research* 13: 197-204.
- Leica Geosystems. 2005. ERDAS field guide. Leica Geosystems Geospatial Imaging, Georgia, USA. LLC. Norcross. 674 p.
- Leites L, AP Robinson. 2004. Improving taper equations with crown dimensions in a mixed-effects modeling framework. *For. Sci.* 50: 204-212.
- LeMay V, H Temesgen. 2005. Comparison of nearest neighbor methods for estimating basal area and stems per hectare using aerial auxiliary variables. *For. Sci.* 51(2): 109-119.
- Lencinas JD. 2002. Stand volume estimation from satellite data employing two-phased sampling. In 29th International Symposium on Remote Sensing of Environment, abril 8-12, 2002. Buenos Aires, Argentina.
- Lencinas JD, D Mohr-Bell. 2007. Estimación de clases de edad de las plantaciones de la provincia de Corrientes, Argentina, con base en datos satelitales Landsat. *Bosque* 28(2): 106-118.
- Lizzi JM, MF Garbulsky, RA Galluscio, AV Deregibus. 2007. Mapeo indirecto de la vegetación de Sierra de la Ventana, provincia de Buenos Aires. *Ecología Austral* 17: 217-230.
- Lu D, P Mausel, E Brondízio, E Moran. 2004. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *For. Ecol. Manage.* 198: 149-167.
- Mäkelä H, A Pekkarinen. 2004. Estimation of forest stand volumes by Landsat TM imagery and stand level field-inventory data. *For. Ecol. Manage.* 196: 245-255.
- Markham BL, G Chander. 2003. Revised Landsat 5 TM Radiometric Calibration Procedures and Post-Calibration Dynamic Ranges. *IEEE Transactions on geoscience and remote sensing* 41: 2674-2677.
- McRoberts RE, MD Nelson, DG Wendt. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the k-nearest neighbors technique. *Remote Sensing of Environment* 82: 457-468.
- McRoberts RE, EO Tomppo. 2007. Remote sensing support for national forest inventories. *Remote Sensing of Environment* 110: 412-419.
- Mehtätalo L. 2004. A longitudinal height-diameter model for Norway spruce in Finland. *Can. J. For. Res.* 34: 131-140.
- Meng Q, CJ Cieszewski, M Madden, B Borders. 2007. A linear mixed-effects model of biomass and volume of trees using Landsat ETM+ images. *For. Ecol. Manage.* 244: 93-101.
- Mery G, C Bahamóndez. 1995. A remote sensing-based forest inventory in the natural forests of Malleco, Chile. Department of Forest Resource Management Publications. Helsinki, Finland. University of Helsinki. 43 p.
- Mevik BH, HR Cederkvist. 2004. Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR). *Journal of Chemometrics* 18(9): 422-429.
- Mevik BH, R Wehrens. 2007. The pls package: Principal component and partial least squares regression in R. *Journal of Statistical Software* 18(2): 1-24.
- Moeur M, AR Stage. 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. *For. Sci.* 41(2): 337-359.
- Montaldo BP. 1974. La bioecología de *Araucaria araucana* (Mol.) Koch. *Boletín Técnico Instituto Forestal Latino-Americano de Investigación y Capacitación* 46/48: 3-55.
- Mujica R. 2001. Untersuchungen zur waldbaulichen Behandlung von *Araucaria araucana* Wäldern in Südchile. Dr. Thesis. München, Deutschland. Fakultät für Ernährung, Technische Universität München. 190 p.
- Navarro RM, M Rosenfeld, J Pérez-Aranda, E Pedrón, JR Guzmán, R Hernández, L González. 2008. Evaluación de la mortalidad de bosques de ñirre (*Nothofagus antarctica*) en la Patagonia chilena mediante imágenes Landsat TM y ETM+. *Bosque* 29(1): 65-73.
- Nothdurft A, J Saborowski, J Breidenbach. 2009. Spatial prediction of forest stand variables. *Eur. J. For. Res.* 128(3): 241-251.
- Ojeda N, Herrera MA. 2000. Clasificación de la vegetación nativa del Parque Nacional Conguillío en la cordillera de los Andes en el centro-sur de Chile, a través de Landsat TM y sistemas de información geográfica. In Institute of Geodesy and Cartography in Warsaw (ed.) Remote Sensing and forest monitoring. Warsaw, Poland. Institute of Geodesy. p. 55-70.
- Packalén P, M Maltamo. 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment* 109(3): 328-341.
- Pinheiro JC, DM Bates. 2000. Mixed-effects Models in S and Spls. New York, USA. Springer-Verlag. 528 p.
- Pinheiro J, D Bates, S DebRoy, D Sarkar, the RCoreTeam. 2008. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-89.
- Pollmann W. 2003. Stand structure and dendroecology of an old-growth *Nothofagus* forest in Conguillío National Park, south Chile. *For. Ecol. Manage.* 176: 87-103.
- Poso S, T Hame, R Paananen. 1984. A method of estimating the stand characteristics of a forest compartment using satellite imagery. *Silva Fennica* 18: 261-292.
- Potvin C, DA Roff. 1993. Distribution-free and robust statistical methods: Viable alternatives to parametric statistics. *Ecology* 74(6): 1617-1628 (Errata, vol.76 (6): 2000).
- Puente M. 1980. Utilización de un bosque del tipo forestal *Araucaria* con criterios de permanencia. 22 p. (Boletín Técnico N° 57).
- R Development Core Team. 2008. R: a language and environment for statistical computing, available from <http://www.R-project.org> [version 2.8.1]. R Foundation for Statistical Computing, Vienna, Austria.
- Rechene C, J Bava, R Mujica. 2003. Los bosques de *Araucaria araucana* en Chile y Argentina. Eschborn, Alemania. Programa de Apoyo Ecológico (TOEB), TWF-40s, GTZ. 158 p.
- Reese H, M Nilsson, P Sandström, H Olsson. 2002. Applications using estimates of forest parameters derived from satellite and forest inventory data. *Computers and Electronics in Agriculture* 37: 37-55.
- Richard J. 1993. Remote Sensing Digital Image Analysis. An Introduction. Berlin, Germany. Springer-Verlag. 333 p.

- Robinson GK. 1991. That BLUP is a good thing: The estimation of random effects. *Statistical Science* 6(6): 15-51.
- Ruiz-Tagle C. 2003. Aplicación de *data minings* sobre un GIS de caracterización forestal. Tesis Ingeniero Civil en Informática. Valdivia, Chile. Facultad de Ciencias de la Ingeniería. Universidad Austral de Chile. 204 p.
- Salas C. 2002. Ajuste y validación de ecuaciones de volumen para un relicto del bosque de roble-laurel lingue. *Bosque* 23(2): 81-92.
- Salas C, L Ene, TG Gregoire, E Næsset, T Gobakken. 2010. Modelling tree diameter from airborne laser scanning derived variables: a comparison of spatial statistical models. *Remote Sensing of Environment* 114(6): 1277-1285.
- Salas C, García O. 2006. Modelling height development of mature *Nothofagus obliqua*. *For. Ecol. Manage.* 229(1-3): 1-6.
- Salas C, V LeMay, P Núñez, P Pacheco, A Espinosa. 2006. Spatial patterns in an old-growth *Nothofagus obliqua* forest in south-central Chile. *For. Ecol. Manage.* 231(1-3): 38-46.
- Salas C, AR Stage, AP Robinson. 2008. Modelling effects of overstory density and competing vegetation on tree height growth. *For. Sci.* 54(1): 107-122.
- Segura R, G Trincado. 2003. Cartografía digital de la reserva nacional Valdivia a partir de imágenes satelitales Landsat TM. *Bosque* 24(2): 43-52.
- Siebert H. 1995. Estudio de crecimiento para un bosque multietáneo de lenga (*Nothofagus pumilio* (Poepp et Endl.) Krasser) en Aysén, XI Región. Tesis Ingeniero Forestal. Santiago, Chile. Universidad de Chile. 75 p.
- Sivanpillai R, MG Messina, X Wu. 2006. Estimation of managed loblolly pine stand age and sensity with Landsat ETM+ data. *For. Ecol. Manage.* 223: 247-254.
- Smith SM. 1995. Comment of: Distribution-free and robust statistical methods: Viable alternatives to parametric statistics? *Ecology* 76(6): 1997-1998.
- Stage AR, NL Crookston. 2007. Partitioning error components for accuracy-assessment of near-neighbor methods of imputation. *For. Sci.* 53: 62-72.
- Stage AR, C Salas. 2007. Interactions of elevation, aspect and slope in models of forest species composition and productivity. *For. Sci.* 53(4): 486-492.
- Stewart-Oaten A. 1995. Rules and judgments in statistics: Three examples. *Ecology* 76(6): 2001-2009.
- Teillet PM, G Fedosejevs. 1995. On the dark target approach to atmospheric correction of remotely sensed data. *Can. Journal of Remote Sensing* 21: 374-387.
- Thompson B. 1995. Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement* 55(4): 525-534.
- Tomppo E. 1991. Satellite image-based national forest inventory of Finland. *International Archives of Photogrammetry and Remote Sensing* 28(7-1): 419-424.
- Tomppo E, M Hakana, M Katila, J Perasaari. 2008. Multi-Source National Forest Inventory. Methods and Applications. Dordrecht, The Netherlands. Kluwer Academic Publishers. 374 p.
- Trotter CM, JR Dymond, CJ Goulding. 1997. Estimation of timber volume in a coniferous forest using Landsat TM. *Int. J. Rem. Sensing* 18: 2209-2223.
- UICN (Unión Internacional para la Conservación de la Naturaleza, CH). 2001. Categorías y criterios de la lista roja de la UICN. Informe técnico 3.1, Unión Internacional para la Conservación de la Naturaleza y de los Recursos Naturales. Gland, Suiza, y Cambridge, Reino Unido. Comisión de Supervivencia de Especies de la UICN. 33 p.
- Ustin SL, A Trabucco. 2000. Using hyperspectral data to assess forest structure. *J. Forestry* 98(6): 47-49.
- Vassilopoulou S, L Hurni, V Dietrich, E Baltsavias, M Pateraki, E Lagios, I Parcharidis. 2002. Orthophoto generation using IKONOS imagery and high-resolution DEM: a case study on volcanic hazard monitoring of Nisyros Island (Greece). *ISPRS Journal of Photogrammetry and Remote Sensing* 57: 24-38.
- Veblen TT. 1982. Regeneration patterns in *Araucaria araucana* forests in Chile. *Journal of Biogeography* 9(1): 11-28.
- Veblen TT, AT Veblen, FM Schlegel. 1979. Understory patterns in mixed evergreen-deciduous *Nothofagus* forests in Chile. *J. Ecol.* 67: 809-823.
- Veblen TT, FM Schlegel. 1982. Reseña ecológica de los bosques del sur de Chile. *Bosque* 4(2): 73-115.
- Veblen TT, C Donoso, T Kitzberger, AJ Rebertus. 1996a. Ecology of southern Chilean and Argentinean *Nothofagus* forests. In Veblen TT, RS Hill, J Read eds. The ecology and biogeography of *Nothofagus* forests. New Haven and London, UK. Yale University Press. p. 293-353.
- Veblen TT, T Kitzberger, B Burns, A Rebertus. 1996b. Perturbaciones y dinámica de regeneración en bosques andinos del sur de Chile y Argentina. In Armesto J, C Villagrán, M Arroyo eds. Ecología de los bosques nativos de Chile. Santiago, Chile. Editorial Universitaria. p. 169-197.
- Vickers AJ. 2005. Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC Medical Research Methodology* 5: 35.
- Waller LA, CA Gotway. 2004. Applied spatial statistics for public health data. Hoboken, N. J, USA. John Wiley. 494 p.
- Weinand DC. 2007. A study of parametric versus non-parametric methods for predicting paleohabitat from Southeast Asian bovid astragali. *Journal of Archaeological Science* 34(11): 1774-1783.
- Wold H. 1975. Path models with latent variables: the NIPALS approach. In Blalock HM ed. Quantitative sociology: International perspectives on mathematical and statistical modeling. New York, USA. Academic Press. p. 307-357.
- Wold S, H Martens, H Wold. 1983. The multivariate calibration problem in chemistry solved by the PLS method. In Ruhe A, Kågström eds. Matrix pencils. Heidelberg. Springer Verlag. p. 286-293.
- Wold S. 2001. Personal memories of the early PLS development. *Chemometrics and Intelligent Laboratory Systems* 58(2): 83-84.

Recibido: 08.04.10

Aceptado: 07.06.10