

## Comparación de redes neuronales y regresión lineal para estimar productividad de sitio en plantaciones forestales, utilizando geomática

Comparison of neural networks and linear regression to estimate site productivity in forest plantations, using geomatic

Carlos Mena Frau<sup>1\*</sup>, Rodrigo Montecinos Guajardo<sup>2</sup>

\*Autor de correspondencia: <sup>1</sup> Universidad de Talca, Departamento de Gestión Forestal Ambiental, Avda. Lircay s/n, Talca, Tel. 56-71-200433, Fax: 56-71-200455, cmena@utalca.cl

<sup>2</sup> Universidad de Talca, Departamento de Gestión Forestal Ambiental, Talca

### SUMMARY

This research paper proposes a methodology to estimate site productivity in forest plantations by both linear regression models and artificial neural networks using some geomatic tools, such as Geographic Information System (GIS), Global Positioning System (GPS) and Fotogrammetry. The study was carried out in the El Picazo Experimental Station which is located in the Andean premountain of the Maule Region. Independent variables are related to the distance to the water courses (DCA), terrain elevation model (MDE), geographic orientation model (MDO) and terrain slope model (MDP); the dependent variable was the total height of the 100 tallest trees per hectare (H\_100). In all the estimation techniques applied the best selected variables were the terrain slope and the distance to the water courses. Moreover, the quality of the generated estimations ( $R^2 = 41.65\%$ ) is within the range established in previous similar research works. The analysis of the results shows that the neural networks has smaller values for absolute mean error (EMA) and mean error quadratic root square (RMSE), than the linear regression model which has a smaller mean bias (SM) and is easier to integrate into a Geographic Information System. Finally, it is possible to establish that the site productivity based on some environmental variables like those indicated above allows to know useful information in order to develop forest plantation programs in bare zones.

*Key words:* geomatic, neural networks, site productivity.

### RESUMEN

En la presente investigación se propone una metodología para estimar la productividad de sitio en plantaciones forestales mediante modelos de regresión lineal y redes neuronales, utilizando herramientas geomáticas, tales como sistemas de información geográfica (SIG), sistema de posicionamiento global (GPS) y fotogrametría. El estudio se llevó a cabo en la Estación Experimental "El Picazo", ubicada en la precordillera andina de la VII Región del Maule. Las variables independientes consideradas tienen relación con la distancia a los cursos de agua (DCA), modelo digital de elevaciones (MDE), modelo digital de orientaciones (MDO) y modelo digital de pendientes (MDP); como variable dependiente se utilizó la altura de los 100 individuos más altos por hectárea (H\_100). En ambas técnicas de estimación, las variables finalmente seleccionadas fueron MDP y DCA. La calidad de las estimaciones generadas ( $R^2 = 41,65\%$ ) se encontró dentro del rango establecido en investigaciones anteriores. El análisis de los resultados establece que el modelo neuronal presenta un menor error medio absoluto (EMA) y una raíz cuadrada del error cuadrático medio (RMSE) inferior respecto al modelo de regresión lineal múltiple, el cual presentó un menor sesgo medio (SM) y fue más fácil de integrar en un SIG. Por último, se destaca que la productividad de sitio basada en indicadores ambientales como los aquí considerados, permite conocer información útil para desarrollar programas de forestación en zonas des pobladas.

*Palabras clave:* geomática, redes neuronales, productividad de sitio.

## INTRODUCCIÓN

La geomática es un campo de actividades que usando una aproximación sistémica integra todos los medios para adquirir y manejar datos espaciales requeridos como parte de las actividades científicas, administrativas, legales y técnicas, que se preocupan de la producción y manejo de información espacial (Instituto Canadiense de Geomática 2000). Se puede afirmar que una de las fortalezas que posee la geomática está referida a la capacidad de combinar información de diferentes fuentes, la cual, resumiendo apropiadamente, es capaz de generar información que en principio no se dispone con facilidad (Lowell 1999).

Por otra parte, la capacidad productiva de un determinado lugar se conoce como calidad de sitio, donde “sitio” está formado por un complejo de factores bióticos y abióticos y su calidad está en función de los factores ambientales relacionados con el suelo, clima, topografía, entre otros (Álvarez y Ruiz 1995). Por lo tanto, la productividad de los sitios forestales puede definirse como el máximo volumen de madera que se puede obtener en un lugar y tiempo determinados. Este parámetro puede ser expresado mediante el modelo general:  $P = f(C, R, S, G, V, A, M, T)$ , donde P: productividad de sitio, C: clima, R: relieve, S: factores asociados al suelo, G: calidad genética, V: estructura de la comunidad vegetal, A: animales, M: influencia humana, y T: tiempo (Gerding y Schlatter 1995).

La relación funcional indicada anteriormente,  $P = f$  (Variables ambientales y tiempo), da origen a modelos de productividad mediante el empleo de modelos de regresión lineal múltiple (Green *et al.* 1989, Klinka y Carter 1990, Rodrigue 1997).

Una “red neuronal artificial” (RNA) está constituida por una colección de elementos de procesamiento (nodos o neuronas) altamente interconectados que transforma un conjunto de datos de entrada en un conjunto de datos de salida deseado (Iost y Rivera 1993, Zhou y Zivco 1996). Es una técnica de estimación y clasificación perteneciente a la inteligencia artificial, que tiene como principio el proceso de aprendizaje que intenta simular la conducta cognitiva del cerebro humano (Freeman y Skapura 1993).

Rath (1999) y Schultz *et al.* (1999) señalan que una RNA está constituida por capas o *layers* de información, donde generalmente se puede distinguir una capa de entrada (variables independientes), una o varias capas intermedias u ocultas (que realizan la determinación de las relaciones entre las variables de entrada y salida) y una capa de salida que recibe el resultante de las variables independientes (figura 1).

El objetivo de la presente investigación fue establecer una metodología para estimar y comparar la productividad de sitio en plantaciones forestales, utilizando técnicas de regresión lineal y redes neuronales, en donde se relacione un índice de productividad (altura dominante

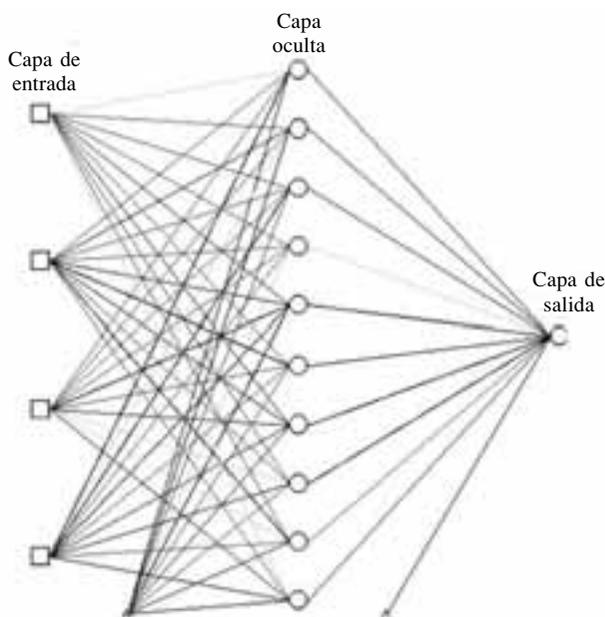


Figura 1. Red neuronal artificial.  
Artificial neural network.

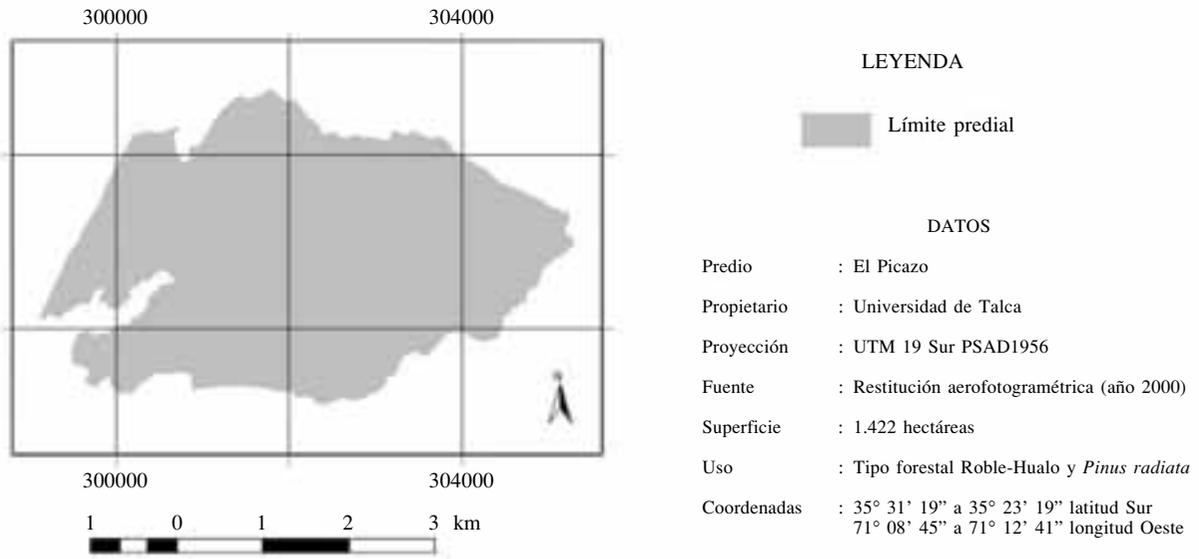
de los árboles a los 6 años de edad) con variables ambientales obtenidas de una base de datos sustentada por un sistema de información geográfica (SIG).

## MÉTODOS

**Área de estudio.** La zona de estudio considerada para la realización de la presente investigación correspondió al predio “El Picazo” perteneciente a la Universidad de Talca. Se encuentra ubicado en la VII Región del Maule, comuna de San Clemente, entre los 35° 31’ 19’’ a 35° 23’ 19’’ de latitud sur y los 71° 08’ 45’’ a 71° 12’ 49’’ de longitud oeste, contando con una superficie de 1.422 ha (figura 2).

**Toma de datos.** Del inventario forestal disponible para la zona considerada se obtuvo la variable dependiente altura total (m) para los 100 árboles más altos por hectárea, contenidos en una parcela circular de 500 metros cuadrados. Se estableció una red de parcelas sistemática con una intensidad de muestreo de una parcela cada tres hectáreas. Se midió un total de 211 parcelas en el predio.

La información cartográfica digital empleada consideró aquellas variables que tuvieran alguna incidencia en el comportamiento de la productividad de sitio en plantaciones de *Pinus radiata* D. Don. Las variables espaciales seleccionadas de la base de datos fueron: red hidrográfica, curvas de nivel (equidistancia de 10 m) y ubicación geográfica de las parcelas. Estas coberturas se encuentran en formato vectorial y provienen de una restitución óptico-digital de fotografías aéreas verticales pancromáticas.



**Figura 2.** Zona de estudio. Predio El Picazo.

Study zone. El Picazo farm.

cas, escala 1:20.000, logrando un nivel de detalle final 1:5.000. A partir de la información espacial anterior se generaron los datos a emplear en los modelos de productividad de sitio, incluyendo modelo digital de elevaciones (MDE), de pendientes (MDP), de orientaciones (MDO) y distancia a los cursos de agua (DCA).

*Generación de los planos digitales.* La generación de la cartografía digital utilizada contempló la realización de las siguientes fases:

- i. *Construcción del modelo digital de elevaciones (MDE).* Una de las formas más comunes de obtener un MDE en formato *raster* es la interpolación lineal de curvas de nivel. Para realizar este proceso, el software IDRISI se vale del máximo valor de la pendiente de los píxeles vecinos para calcular el valor de la elevación. La validación del MDE fue realizada mediante una comparación de las mediciones realizadas en terreno con un receptor GPS cartográfico y las estimadas por el modelo. Este procedimiento arrojó un RMS (*Root Mean Square*) de 2,79 m.
- ii. *Construcción de los modelos derivados.* Para la construcción del modelo digital de pendientes (MDP) el sistema determina el máximo valor de las pendientes incluidas en una ventana móvil de 3x3, asignándolo a la celda central. Para el modelo digital de orientaciones (MDO), las exposiciones del terreno se calcularon en grados sexagesimales y en el sentido de las manecillas del reloj. A los sectores planos el algoritmo le asigna el valor temático -1.

iii. *Distancia a los cursos de agua (DCA).* La variable se generó en el software *raster*, donde el valor de cada celda se calculó como la distancia euclidiana entre dicho píxel y los de referencia.

La información cartográfica mencionada anteriormente servirá de base para la construcción de los modelos de productividad de sitio mediante técnicas de regresión lineal y redes neuronales. Esta información se obtuvo mediante la realización de un proceso de restitución fotogramétrica óptico-digital, que de acuerdo a lo expuesto por López y Mena (1999) es la técnica que presenta una mayor confiabilidad para la generación y actualización cartográfica.

*Generación de los modelos de productividad de sitio mediante regresión lineal múltiple.* A partir del método mínimos cuadrados ordinarios, se generaron curvas que explican el comportamiento de la productividad de sitio en plantaciones forestales. Para la construcción del modelo se utilizó un 70% de los datos disponibles y para su validación el 30% restante. Es conveniente señalar que la gran mayoría de los análisis realizados se basan en el uso de un estadístico denominado valor-p o *P-Value*, el cual corresponde a la probabilidad de aceptar la hipótesis nula, comparada con el nivel de significancia  $\alpha$  (se utilizó  $\alpha = 0,05$ ). A continuación se presenta la secuencia lógica para efectuar el análisis de regresión lineal múltiple:

- i. *Selección de variables.* Debido a que no se conoce cuáles son las variables influyentes en la productivi-

dad de sitio, se utilizó una matriz de correlaciones basada en el cálculo del coeficiente de correlación de Pearson, para establecer una primera aproximación sobre la determinación de variables independientes (Vallejos 1999). La selección final de las variables explicativas del modelo se realizó empleando el algoritmo *Forward Selection* implementado en el software *Statgraphics Plus*.

- ii. *Análisis de regresión*. En esta fase se determinó el valor de los coeficientes e intercepción de la recta de regresión. Además, se obtuvieron los estadísticos básicos para analizar el modelo: coeficiente de determinación ( $R^2$ ), coeficiente de determinación ajustado ( $R_a^2$ ), el error estándar de la estimación y el aporte de cada variable en el modelo (Canavos 1988).
- iii. *Supuestos acerca del error*. Anderson *et al.* (1999) señalan que es necesario verificar tres supuestos básicos contemplados en el método de mínimos cuadrados ordinarios: normalidad, media igual a cero y homocedasticidad. Normalidad, mediante la prueba Kolmogorov-Smirnov (K-S), apropiada para el estudio de variables aleatorias continuas; este análisis se complementó mediante el gráfico de probabilidad normal y el histograma de frecuencias de los errores (Cook y Weisberg 1999). Media igual a cero: se probó el supuesto que evalúa si el valor medio de los residuos es igual a cero (Vallejos 1999). Para ello, se utilizó la siguiente prueba de hipótesis:  
H<sub>0</sub>: la media de las observaciones es igual a cero.  
H<sub>a</sub>: la media de las observaciones es distinta de cero.  
Homocedasticidad: este supuesto sugiere que las varianzas de la variable independiente sean similares. Para determinarlo se utilizó la prueba de Bartlett (Gujarati 1996).

*Generación de modelos de productividad de sitio mediante redes neuronales artificiales*. A partir del algoritmo de redes neuronales artificiales se establecieron los modelos que explican el comportamiento de la productividad de sitio en plantaciones de *Pinus radiata*, utilizando variables topográficas y ambientales. Para tales efectos se utilizó el software *Pathfinder Neural Networks System™*. El ingreso de los datos al modelo requirió de una normalización, vale decir, los valores fueron escalados entre cero y uno. La construcción y validación del modelo neuronal se llevó a cabo mediante la realización de las siguientes etapas:

- i. *Fase de entrenamiento*. La fase contempló la utilización del 45% de los datos, con la finalidad de calcular los pesos sinápticos de las conexiones. Se probaron varias arquitecturas de la red neuronal, utilizando sólo una capa oculta o intermedia dentro de la configuración, ya que el software usado no permite ampliar la cantidad de *layers*. El algoritmo empleado para el proceso de entrenamiento fue el de

retropropagación del error, considerando una función de transferencia sigmoidea. Para la selección de las variables que explican el modelo y el reconocimiento de los puntos de quiebre, se realizó la gráfica del RMSE y el coeficiente de correlación de Pearson, cuantificando el aporte de cada variable al modelo. Para evitar el riesgo de sobreentrenamiento se adoptó un criterio de detención basado en el análisis de la gráfica del RMSE de la curva de control y de entrenamiento.

- ii. *Fase de prueba*. Se utilizó un 10% de los datos restantes, con el objeto de determinar el momento oportuno cuando el proceso de entrenamiento puede detenerse, definiendo posteriormente la matriz final de pesos sinápticos.
- iii. *Fase de validación*. La validación del modelo neuronal se llevó a cabo empleando el 15% restante de los datos y una vez que el entrenamiento ha finalizado. Su objetivo es medir la calidad del modelo que se ha propuesto y verificar si fue capaz de generalizar los datos en el proceso de aprendizaje. Como estadísticos de validación se utilizaron el RMSE y el coeficiente de correlación de Pearson.

*Comparación de los modelos de productividad de sitio*. Los modelos de productividad generados fueron validados con aquellos datos no empleados para su construcción, utilizando para esta tarea el 30% de la información disponible. Los índices considerados para la comparación de los modelos fueron el RMSE, el EMA (error medio absoluto) y el SM (sesgo medio), este último usado para analizar el sentido de las desviaciones (sobre o subestimación).

*Aplicación de los modelos de productividad de sitio*. Posterior a la evaluación y validación de los modelos, éstos se implementaron al interior de un sistema de información geográfica (IDRISI), obteniéndose mapas temáticos de productividad de sitio para plantaciones de *Pinus radiata* presentes en la zona de estudio.

## RESULTADOS

*Estimación por el método de mínimos cuadrados*. A continuación se describen los resultados obtenidos de los modelos de regresión lineal múltiple:

*Análisis de correlación*. En el cuadro 1 se presentan las correlaciones de Pearson con respecto a la variable dependiente (productividad de sitio), representada por la altura promedio de los 100 árboles más altos por hectárea a los 6 años de edad (H<sub>100</sub>). Ello representa el análisis exploratorio de las variables independientes con el objeto de observar su influencia sobre la productividad de sitio.

**Cuadro 1.** Análisis de correlación.  
Correlation analysis.

Variable	Coeficiente de correlación (Valor-P)	
MDE (Modelo digital de elevaciones)	0,1452	(0,0804)
MDO (Modelo digital de orientaciones)	-0,1198	(0,1499)
MDP (Modelo digital de pendientes)	-0,4849	(0,0000)
DCA (Distancia a cursos de agua)	-0,4539	(0,0000)

Una vez establecidas las variables que presentan un mayor coeficiente de correlación de Pearson (MDP y DCA) en relación con la productividad de sitio ( $H_{100}$ ) se procede, mediante el algoritmo *Forward Selection*, a determinar las variables independientes y sus transformaciones más importantes que aportan al modelo de productividad de sitio. En este caso los mejores resultados corresponden a  $MDP^2$  y DCA.

*Estimación de los parámetros de regresión.* Determinadas las variables que tienen mayor incidencia en la productividad de sitio, se obtuvieron los parámetros de regresión, coeficientes de determinación y el error estándar de la estimación (cuadro 2).

En el modelo propuesto, tanto la intercepción como las variables aportaron al modelo de regresión (ver valor-P en cuadro 2). En consecuencia, la ecuación quedó definida por la expresión:

$$H_{100} = 8,01713 - 0,00044MDP^2 - 0,00632DCA \quad [1]$$

Es necesario consignar que el valor T de la intercepción es muy elevado y, por lo tanto, la constante (8,01713) explica gran parte de la variación presente en la variable  $H_{100}$ . Por otra parte, los coeficientes de regresión tienen una incidencia baja dentro del modelo, aun cuando aportan estadísticamente.

**Cuadro 2.** Estimación de los parámetros de regresión.  
Estimation of the regression parameters.

Variable Dependiente: Altura de los 100 árboles superiores por hectárea ( $H_{100}$ )				
Parámetro	Estimación	Error estándar	T	Valor-P
Intercepción	8,01713	0,09519	84,2228	0,0000
$MP^2$	-0,00044	0,00006	-7,23018	0,0000
DCA	-0,00632	0,00112	-5,65323	0,0000

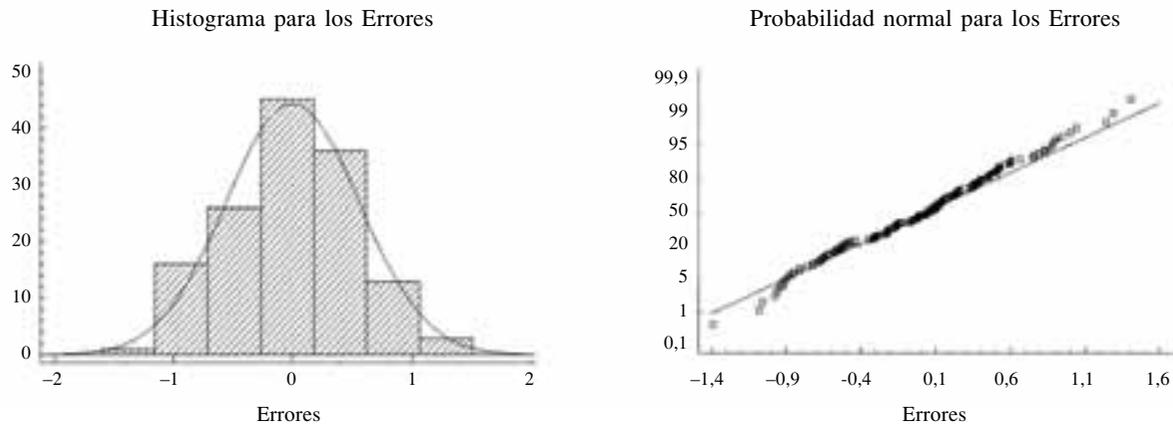
$R^2$ : 41,65%     $Ra^2$ : 40,80%    Error estándar de la estimación: 0,563

*Supuestos acerca del error.* A continuación se muestra el análisis detallado de los errores, paso necesario para el cumplimiento de los supuestos establecidos en el método de mínimos cuadrados.

- Distribución normal de los residuos: con la información entregada por la prueba K-S ( $P = 0,8871$ ) se puede aseverar que los residuos se distribuyeron en forma normal, debido a que el valor-p es mayor que  $\alpha$ , lo cual se expone gráficamente en la figura 3.
- Media de los residuos igual a cero: los resultados entregados por la prueba de hipótesis para la media ( $P = 0,953$ ) indican que no se rechaza la hipótesis nula para un  $\alpha = 0,05$ , razón por la cual es posible aseverar que los residuos tienen media igual a cero.
- Homocedasticidad: considerando la información suministrada por la prueba de Bartlett, donde el valor-P (0,245) es superior al nivel de significancia a (0,05), queda de manifiesto que no se puede rechazar la hipótesis nula; en consecuencia es posible afirmar que los residuos son homocedásticos.

#### ESTIMACIÓN DE LA PRODUCTIVIDAD DE SITIO POR REDES NEURONALES ARTIFICIALES

*Determinación de las variables de entrada en el modelo.* En el cuadro 3 se presenta un resumen con el aporte de cada variable, considerando una estructura arbitraria 3:6:1 (tres nodos de entrada, seis intermedios y uno de salida). El análisis conjunto del RMSE y el coeficiente de correlación permite observar que al extraer la variable MDP y DCA se genera un aumento del RMSE (0,0663 y 0,0597, respectivamente), y una disminución del coeficiente de correlación (0,2606 y 0,4053). Lo indicado anteriormente permite establecer que dichas variables son las mejores predictoras para el modelo neuronal.



**Figura 3.** Histograma de frecuencias y gráfico de probabilidad normal.  
 Frequency histogram and normal probability plot.

**Cuadro 3.** Aporte de cada variable al modelo neuronal.  
 Contribution of each variable to the neural model.

Variable a extraer del modelo	Correlación (Valor-P)	RMSE
MDP	0,2606 (0,0040)	0,0663
DCA	0,4053 (0,0003)	0,0597
MDT	0,5205 (0,0001)	0,0589
MDO	0,6035 (0,0001)	0,0518

*Definición del número de nodos ocultos.* Con el conocimiento de las variables significativas se determinó la arquitectura final del modelo neuronal, observando en forma simultánea el coeficiente de correlación y RMSE. El cuadro 4 presenta un resumen de siete estructuras evaluadas para el modelo neuronal. La cantidad de iteraciones varió dependiendo de la estructura seleccionada, el criterio de parada está relacionado con el análisis de la gráfica del RMSE en función del tiempo de la curva de control y de aprendizaje para cada arquitectura del modelo neuronal.

**Cuadro 4.** Indicadores de la calidad del modelo neuronal con diferentes estructuras.  
 Neural model quality indicators with different structures.

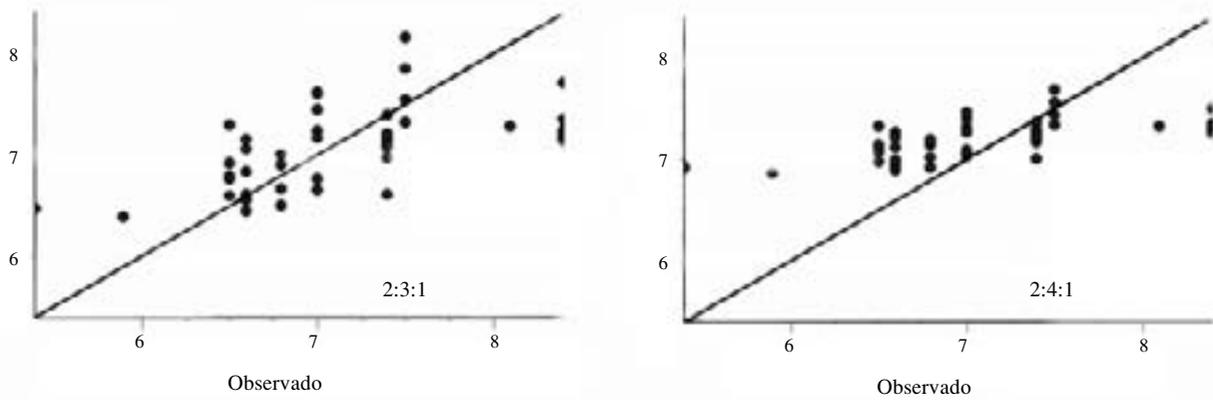
Cantidad de nodos ocultos	Coefficiente de Correlación (Valor-P)	RMSE
2	0,5362 (0,0002)	0,0527
3	0,5760 (0,0001)	0,0504
4	0,5923 (0,0000)	0,0539
5	0,5310 (0,0002)	0,0529
6	0,5468 (0,0002)	0,0521
7	0,5760 (0,0001)	0,0506
8	0,5405 (0,0002)	0,0523

De acuerdo a la información obtenida del cuadro 4, se concluye que la mejor estructura es la 2:3:1, si se considera el RMSE como estimador de la calidad. Por otro lado, si se considera el coeficiente de correlación, la mejor estructura es la 2:4:1. Por este motivo, se decidió recurrir al análisis gráfico para escoger la mejor estructura. Este análisis (figura 4) permite apreciar una mejor distribución de los datos en el primer caso (estructura 2:3:1), ya que los valores reales se disponen de mejor manera sobre la recta del valor estimado, en comparación con los resultados suministrados por el modelo neuronal de estructura 2:4:1 ubicado a la derecha.

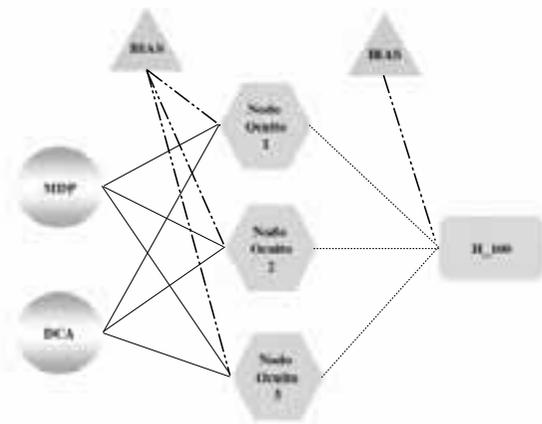
La figura 5 muestra la arquitectura definitiva de la red neuronal seleccionada. En la capa de entrada se observa la presencia del modelo digital de pendientes (MDP) y la distancia a los cursos de agua (DCA). En la capa intermedia se simbolizan tres nodos ocultos y en la de salida se encuentra la altura de los 100 árboles más altos por hectárea (H\_100). Finalmente, se observa la presencia de los BIAS, correspondientes a valores numéricos que actúan como constantes independientes de la ponderación de las variables de entrada.

El algoritmo de retropropagación del error (*Backpropagation*) con tres neuronas ocultas y 6.810 ciclos realizados en el *software Pathfinder Neural Networks System™* fue la estructura seleccionada para la construcción del modelo final.

*Comparación de los modelos de productividad de sitio.* El cuadro 5 presenta los resultados del análisis de la comparación de los modelos obtenidos por técnicas de regresión lineal múltiple y redes neuronales artificiales. Se aprecia que el modelo neuronal presenta un menor error asociado a la estimación cuando se considera al RMSE y EMA como parámetros de comparación. Por otra parte, si se contempla el sesgo medio como indicador, el modelo de regresión lineal presenta un menor error asociado en sus estimaciones.



**Figura 4.** Valores observados en función del valor calculado.  
Actual values in relationship with calculate value.



**Figura 5.** Estructura de la red neuronal.  
Neural network structure.

*Aplicación de los modelos de productividad de sitio.* Finalmente, los modelos de regresión lineal y de redes neuronales escogidos se implementaron en el software IDRISI, obteniéndose un mapa de productividad de sitio del cual se presentan las categorías y superficies asociadas en el cuadro 6.

De acuerdo a la información suministrada por el cuadro 6 se puede afirmar que existe una concentración de superficies en las categorías correspondientes entre los 6 y 8 metros de altura de los 100 árboles más altos por hectárea a los seis años de establecidos, aspecto que se puede visualizar en la figura 6.

**DISCUSIÓN**

Como se puede observar en el cuadro 6, el modelo neuronal generaliza la información de productividad de sitio, mientras que el modelo resultante de la regresión

**Cuadro 5.** Comparación de los modelos de productividad de sitio.  
Site productivity models comparisons.

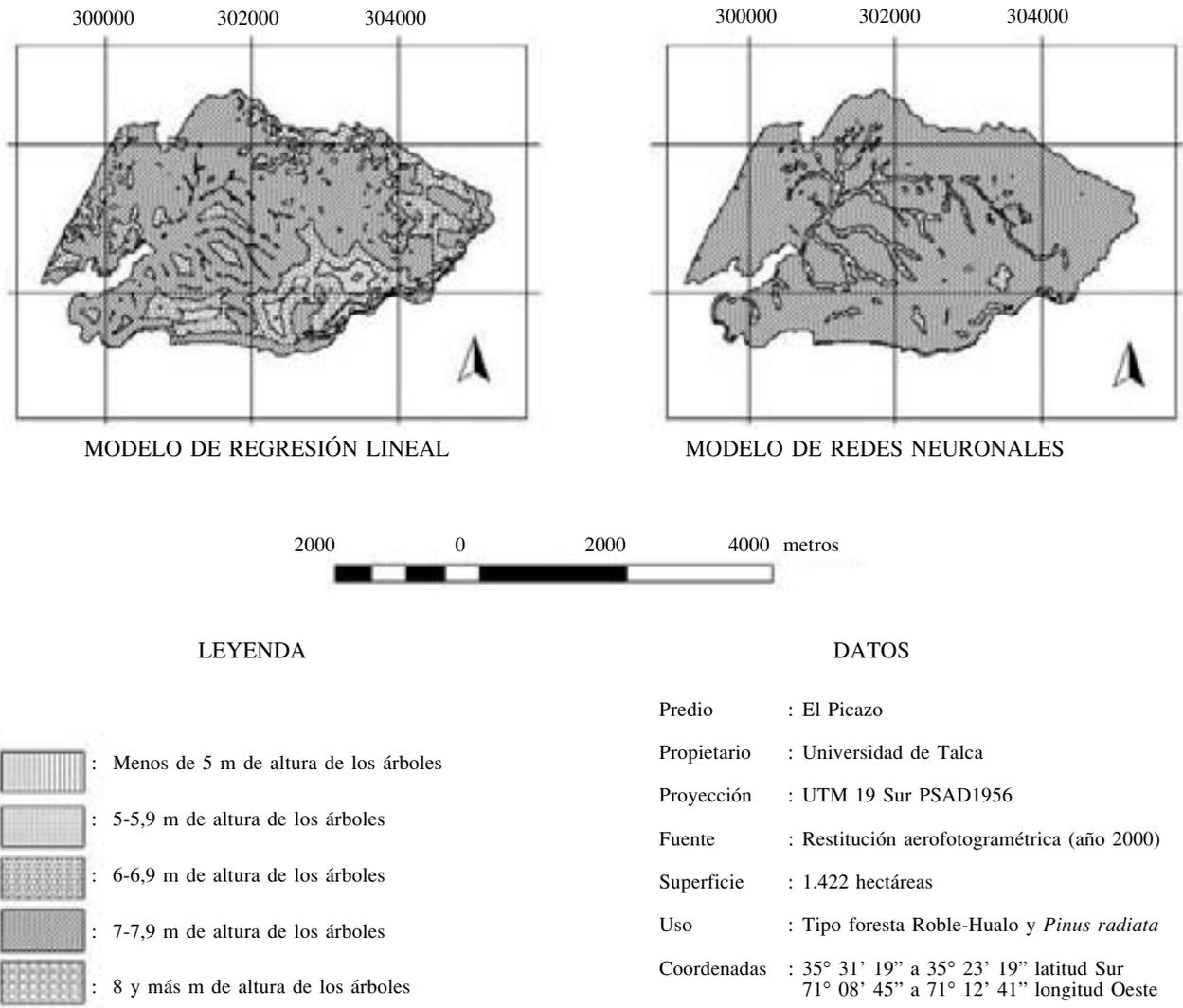
Modelo	Índice		
	SM	EMA	RMSE
Regresión lineal	-0,0159	0,4551	0,5619
Redes neuronales	0,0382	0,4380	0,5548

**Cuadro 6.** Superficies en hectáreas por categoría de alturas.  
Land area of each category, in hectares.

Categoría (H_100)	Modelo de regresión lineal	Modelo de redes neuronales
0-5	8,74	-
5-6	58,26	-
6-7	389,42	18,55
7-8	957,55	1.278,71
8 y más	8,76	125,47

lineal genera una mayor diversidad de situaciones de productividad para *Pinus radiata* creciendo en la precordillera andina de la VII Región del Maule. La representación digital de los modelos de productividad de sitio permite clasificar el territorio en zonas de diferente calidad (figura 6).

Por otro lado, analizando los resultados queda de manifiesto que el modelo neuronal presenta un menor error medio absoluto (EMA) y un menor RMSE; por el contrario, el modelo de regresión lineal presenta un menor sesgo medio al realizar la estimación. Aun cuando las variables consideradas no explican del todo el com-



**Figura 6.** Cartas temáticas de productividad de sitio.  
 Thematic chart of site productivity.

portamiento de la productividad del sitio, se encuentran dentro de los rangos establecidos por otros estudios similares (De las Salas 1984, Hairston y Grigal 1991, citados por Vallejos 1999). En vista de ello, Daniel *et al.* (1982) señalan que es altamente recomendable para estudios futuros considerar variables asociadas al suelo (estructura, textura, materia orgánica, profundidad, contenido de humedad, pH, entre otras) y clima (precipitación, temperatura, humedad relativa, entre otras), que puedan representarse espacialmente en una base de datos digital sustentada por un sistema de información geográfica. En este contexto, Schlatter y Gerding (1995) desarrollaron un sistema de clasificación de sitios forestales que se fundamenta en factores según su grado de incidencia en la productividad, ubicando en primer lugar el macroclima, seguido del clima local y las propiedades físicas del suelo, y finalmente las propiedades químicas y nutritivas. Con esto, se efectuó una subdivisión en zonas de

crecimiento (clima), distritos de crecimiento (clima) y áreas de crecimiento (clima y suelo).

Por otra parte, se debe tener especial cuidado en el proceso de aprendizaje o entrenamiento de la red neuronal en la construcción de un modelo utilizando el algoritmo *Backpropagation* o de retropropagación del error, especialmente en lo referido al riesgo de sobreentrenamiento. Para ello es de vital importancia asignar parámetros adecuados a la red en este proceso, además de efectuar el análisis de la gráfica del RMSE en función del tiempo de la curva de aprendizaje y control, de manera de establecer el momento propicio para detener el proceso de aprendizaje y evitar con esto el riesgo de sobreentrenar la red. Este último aspecto es esencial para lograr una buena confiabilidad de las estimaciones efectuadas por el modelo, ya que un sobreentrenamiento de la red neuronal ocasiona que el modelo elaborado no tenga la capacidad de generalizar la información.

## CONCLUSIONES

El empleo de modelos de redes neuronales para representar la productividad de sitio forestal es razonablemente apropiado si se compara con las técnicas tradicionales de regresión lineal múltiple.

Las variables independientes elegidas, a saber: distancia a los cursos de agua y modelo digital de pendientes, explican en parte el comportamiento de la variable altura de los árboles dominantes, utilizada como estimadora de la productividad de sitio.

El modelo de regresión calculado presenta significación estadística de 41,65%, encontrándose dentro de los rangos establecidos en estudios similares.

El modelo de red neuronal calculado presentó una estructura de 2 nodos de entrada, 3 nodos ocultos y 1 de salida, para modelar la productividad de sitio en las plantaciones de *Pinus radiata* presentes en la zona de estudio. Este modelo realiza estimaciones con un menor error asociado a ellas, utilizando el algoritmo de retropropagación del error con el que se debe tener especial cuidado para evitar un sobreentrenamiento de la red.

Por último, para mejorar la significación de los modelos predictivos de calidad de sitio es recomendable incorporar variables independientes asociadas a clima y suelo.

## REFERENCIAS

- Álvarez J, A Ruiz. 1995. *Dasometría: Introducción a las Técnicas de Modelización Forestal*. Madrid, España, Unipia. 83 p.
- Anderson, D, D Sweeney, T Williams. 1999. *Estadística para Administración y Economía*. Ciudad de México, México, International Thomson Editores. 909 p.
- Cook, D, S Weisberg. 1999. *Applied Regression Including Computing and Graphics*. New York, USA, John Wiley & Sons, 593 p.
- Daniel T, J Helms, F Backer. 1982. *Principios de Silvicultura*. New York, USA, Mc Graw Hill. 492 p.
- Canavos, G. 1988. *Probabilidad y Estadística: Aplicaciones y métodos*. Madrid, España, McGraw-Hill. 651 p.
- Freeman J, D Skapura. 1993. *Redes Neuronales: Algoritmos, aplicaciones y técnicas de programación*. New York, USA, Addison-Wesley Iberoamericana. 430 p.
- Gerding V, JE Schlatter. 1995. Variables y factores del sitio de importancia para la productividad de *Pinus radiata* D. Don en Chile. *Bosque* 16(2): 39-56.
- Green R, P Marshall, K Klinka. 1989. Estimating Site Index of Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco) from ecological variables in southwestern British Columbia. 1989. *Forest Science* 35(1): 50-63.
- Gujarati D. 1996. *Econometría*. Ciudad de México, México, McGraw-Hill. 597 p.
- Instituto Canadiense de Geomática. 2000. Definición de Geomática. Consultado 15 mar. 2005. Disponible en <http://www.cartesia.org/article.php?sid=83>.
- Iost C, R Rivera. 1993. Diseño, Implementación y Prueba de una Herramienta Computacional para el manejo de redes neuronales, basada en el mecanismo del Back-Propagation y sus variaciones. Trabajo de título Ingeniero Civil Industrial mención Informática. Temuco, Chile. Universidad de la Frontera. 120 p.
- Klinka K, R Carter. 1990. Relationship Between Site Index and synoptic Environmental Factors in immature Coastal Douglas-Fir Stands. *Forest Science* 36(3): 815-830.
- López CM, CA Mena. 1999. Eliminación del Error de superficie mediante integración de Fotografías aéreas y Sistemas de Información Geográfica y Sistemas de Posicionamiento Global, para evaluar efectos en la Planificación de faenas de inventarios y cosecha forestal. In Mena C ed. *Sistemas de Información Geográfica y Teledetección Espacial aplicados a la Ordenación del Territorio y el Medio Ambiente*. Talca, Chile, Servicio Aerofotogramétrico de la Fuerza Aérea de Chile. p. 117-121.
- Lowell K. 1999. Estimating Forest Site Productivity in an Australian Eucalyptus Forest Using Geomatics. In XIII Silvotecnica. Concepción, Chile. p. 1-10.
- Rath C. 1999. Estimación de la superficie y número de árboles en bosques de *Pinus radiata* D. Don en fotografías aéreas, mediante el uso de redes neuronales artificiales. Memoria Ingeniero Forestal. Santiago, Chile. Facultad de Ciencias Forestales, Universidad de Chile. 108 p.
- Rodrigue J. 2001. Woody species diversity, forest and site productivity, stumpage value, and carbon sequestration of forests on mined lands reclaimed prior to the passage of the surface mining control and reclamation act of 1977. Tesis de Master of Science of Forestry. Virginia, USA. Faculty of the Virginia Polytechnic Institute and State University College of Natural Resources Department of Forestry. 299 p.
- Schlatter, JE, V Gerding. 1995. Método de clasificación de sitios para la producción forestal, ejemplo en Chile. *Bosque* 16(2): 13-20.
- Schultz E, T Metney, J Koger. 1999. A neural network model for wood chip thickness distributions. *Wood and Fiber Science* 31(1): 2-14.
- Vallejos O. 1999. *Dasometría para Ingenieros Forestales*. Talca, Chile, Universidad de Talca. 195 p.
- Zhou J, D Civco. 1996. Using genetic learning neural networks for spatial decision making in GIS. *Photogrammetric Engineering & Remote Sensing* 62(11): 1287-1295.

Recibido: 14.04.05

Aceptado: 09.03.06