



Bosque

ISSN: 0304-8799

revistabosque@uach.cl

Universidad Austral de Chile

Chile

MIRANDA-SALAS, MARCELO; CONDAL, ALFONSO R.

Importancia del análisis estadístico exploratorio en el proceso de interpolación espacial: caso de estudio Reserva Forestal Valdivia

Bosque, vol. 24, núm. 2, agosto, 2003, pp. 29-42

Universidad Austral de Chile

Valdivia, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=173114405004>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica





Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto



## Bosque (Valdivia)

ISSN 0717-9200 *versión on-line*

-  Como citar este artículo
-  Agregar a favoritos
-  Enviar a e-mail
-  Imprimir HTML

Bosque (Valdivia) v.24 n.2 Valdivia ago. 2003

Bosque, Vol. 24 N° 2, 2003, pp. 29-42

### ARTICULOS

## Importancia del análisis estadístico exploratorio en el proceso de interpolación espacial: caso de estudio Reserva Forestal Valdivia

Importance of the exploratory statistics analysis on spacial interpolation process: Study case of Valdivian Forest Reserve

MARCELO MIRANDA-SALAS <sup>1, \*</sup>, ALFONSO R. CONDAL <sup>2</sup>

<sup>1</sup> Departamento de Cs. Forestales, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul, Santiago, Chile. E-mail: [mmirands@puc.cl](mailto:mmirands@puc.cl)

<sup>2</sup> Département des Sciences Géomatiques, Université Laval, Québec, Canada, G1K 7P4 . E-mail: [alfonso.condal@scg.ulaval.ca](mailto:alfonso.condal@scg.ulaval.ca).

\* Actualmente en: Département des Sciences du Bois et de la Forêt, Université Laval, Canada, G1K 7P4.

### Summary

This paper discusses the importance of the exploratory statistical analysis in the interpolation process of spatial data. Basic methods and the hypothesis upon which they are based are reviewed. To illustrate their usefulness, a case study based on forestry data is presented in which information on the

topographic and basal areas of trees is analyzed from a geostatistical point of view. Firstly, it is shown that the set of topographic data used in this work satisfied the requirements for the intrinsic hypothesis of the regionalized variable theory. Secondly, evidence is presented that indicates information on the basal area of trees does not satisfy these conditions. This is attributable to the total lack of spatial correlation shown by these data. Given this evidence, it is finally argued that a successful interpolation procedure cannot be accomplished without first considering the results obtained from an initial or preliminary statistical analysis of the data to be interpolated.

*Key words:* interpolation, geostatistics, digital elevation model, basal area.

## Resumen

En este artículo se discute sobre la importancia del análisis estadístico exploratorio en el proceso de interpolación de datos espaciales. Para esto se han definido las hipótesis básicas de interpolación y seleccionado un conjunto de métodos estadísticos que pueden ser usados en su comprobación. De manera de mostrar su aplicación práctica, un estudio de caso, basado en datos forestales, ha sido planteado utilizando los atributos altitud y área basal de árboles, los cuales han sido analizados desde un punto de vista geoestadístico. Primero, los resultados mostraron que el atributo altitud satisface las condiciones planteadas por la hipótesis intrínseca de variable regionalizada. Segundo, se presentan evidencias que indican que el área basal no satisface dichas condiciones, lo que es atribuible a la ausencia total de correlación espacial en los datos estudiados. Dadas estas evidencias, finalmente se puede indicar que un proceso de interpolación no puede ser exitoso, si primero no considera los resultados obtenidos de un análisis estadístico preliminar de datos a interpolar.

*Palabras claves:* interpolación, geoestadística, modelo digital de terreno, área basal.

---

## INTRODUCCION

La información espacial, en forma de coberturas temáticas asociadas a un Sistema de Información Geográfica (SIG) e imágenes provenientes de sensores remotos, es usada en forma creciente en la modelación de sistemas ecológicos ([Kyriakidis y Dungan 2001](#)). Los datos topográficos, como los valores de las curvas de nivel, son utilizados frecuentemente para generar modelos que representan diferentes características del terreno como altura, pendientes, exposiciones, visibilidad, entre otros, mientras que los datos dasométricos son utilizados para la estimación de la productividad y las existencias de productos ([Burrough y McDonnell 1998](#), [Corvalán et al. 2000](#)).

En este proceso de análisis comúnmente se emplean algoritmos de interpolación espacial que están implementados en programas asociados a Sistemas de Información Geográfica (SIG) de uso comercial, permitiendo crear superficies continuas que pueden tener resoluciones espaciales variables.

La exactitud de los resultados obtenidos por interpolación está asociada principalmente a las estrategias de muestreo y al cumplimiento de las hipótesis estadísticas de los modelos utilizados, en especial aquellas relacionadas con la estructura espacial de los atributos en estudio. Estos aspectos comúnmente son tratados a través del análisis estadístico exploratorio, cuyo objetivo es describir las características globales del conjunto de datos y detectar patrones de regularidad espacial ([Kitanidis 1997](#)).

El análisis exploratorio permite detectar problemas de representatividad en el muestreo, describir la existencia de dependencia espacial en los datos, ajustar un modelo de representación de esta dependencia y estimar el error asociado antes de aplicar un algoritmo de interpolación espacial. Para ello se han desarrollado diversas herramientas como las descripciones estadísticas básicas del ámbito univariado, análisis gráficos y descripciones espaciales basadas en el estudio de la covarianza, semivarianza y autocorrelación espacial ([Isaaks y Srivastava 1989](#), [Goovaerts 1997](#), [Kitanidis 1997](#)). Es importante destacar que la realización de una interpolación espacial, con un análisis estadístico exploratorio previo, permite asegurar la calidad de los resultados obtenidos.

El objetivo de este trabajo es presentar una discusión sobre el uso de herramientas de análisis estadístico exploratorio en el contexto de la interpolación espacial de datos. Para mostrar su aplicación se hace un análisis comparativo de los atributos altitud del terreno y área basal de árboles.

Finalmente, los resultados obtenidos en este estudio pueden ser aplicados a otros atributos espaciales como propiedades de suelos, altura y volumen de bosques, entre otros.

*Interpolación espacial:* La interpolación espacial es un procedimiento matemático utilizado para predecir el valor de un atributo en una localidad precisa a partir de valores obtenidos de puntos vecinos, ubicados al interior de la misma área de estudio. La interpolación se utiliza para transformar un número finito de observaciones, por ejemplo cotas de terreno, en un espacio continuo de manera que su patrón espacial sea comparable con aquel presentado por las observaciones puntuales de base ([Burrough y McDonnell 1998](#)).

La literatura describe diferentes métodos de interpolación. En forma general, ellos son clasificados en globales y locales, de acuerdo a la cantidad de información que utilizan, y en exactos e inexactos, de acuerdo al valor de las estimaciones respecto a los datos de origen ([cuadro 1](#)), ([Burrough y McDonnell 1998](#), [Lam 1983](#)).

CUADRO 1

Clasificación de diferentes métodos de interpolación espacial ([Lam 1983](#), [Burrough y McDonnell 1998](#)).  
Types of spatial interpolation methods ([Lam 1983](#), [Burrough and McDonnell 1998](#)).

Método	Global/Local	Tipo de interpolación
Clasificación	Global	Inexacta
Superficies de tendencia	Global	Inexacta
Modelos de regresión	Global con refinamiento local	Inexacta
Polígonos de Thiessen	Local	Exacta
Interpolación lineal	Local	Exacta
Inverso de la distancia y medias móviles	Local	Inexacta cuando se aplica un suavizamiento de la superficie interpolada
Splines	Local	Exacta, con un refinamiento limitado
Kriging	Local, pero con un variograma global	Exacta
Simulación condicional	Local, pero con un variograma global	Inexacta

Los métodos globales se caracterizan porque las estimaciones son realizadas utilizando todos los datos disponibles en el área de interés, en cambio los métodos locales operan dentro de una pequeña vecindad, alrededor de la ubicación donde se desea obtener un valor interpolado. Por otra parte, los métodos exactos se diferencian de los inexactos debido a que en los primeros el valor interpolado en una localidad donde existe un dato de muestreo no presenta diferencia respecto al valor utilizado de base para la interpolación.

El resultado de un proceso de interpolación espacial va a depender principalmente de las características naturales del atributo en estudio, de la existencia de una dependencia espacial y de factores asociados a la forma de modelización del atributo, dada por la distribución de los datos sobre el terreno, la medición del valor medio, su variabilidad y la elección del modelo predictor ([Lam 1983](#)).

*Variables regionalizadas e interpolación:* En las ciencias forestales, y en la mayoría de las ciencias relacionadas con los recursos naturales, se puede observar que el valor de un atributo muchas veces es dependiente de su localización espacial. Así, el diámetro de un árbol depende directamente de las

características de sus vecinos ([García 1991](#)), la altitud en un punto en el terreno está relacionada con la cota de los puntos de su entorno ([Felicísimo 1994](#)) y la presencia de un nutriente en el suelo no es producto del azar ([Oliver 1990](#)). Estos atributos son llamados variables regionalizadas ya que pueden ser modelados por medio de una función aleatoria que representa su variación natural sobre el espacio a una escala determinada.

De acuerdo a [Isaaks y Srivastava \(1989\)](#), si  $Z(x)$  es el valor de la variable aleatoria  $Z$  en la localidad  $x$ , él puede ser expresado como (ecuación 1):

$$Z(x) = m(x) + \mathcal{E}(x) + \mathcal{E}'' \quad [1]$$

Donde:

$m(x)$  es una función determinística que describe el componente estructural de  $Z(x)$  en la localidad  $x$ ,  $\mathcal{E}(x)$  es un componente que describe la variabilidad estocástica local, pero espacialmente dependiente de  $m(x)$  (llamada la variable regionalizada), y  $\mathcal{E}''$  es un residual o ruido espacialmente independiente con distribución normal.

Para cumplir con la condición de estacionalidad, es decir, la continuidad de las funciones  $m(x)$  y  $e(x)$  sobre el espacio, la teoría de variable regionalizada asume que el promedio de las diferencias entre los valores del atributo  $Z$  en dos localidades  $x$  y  $x+h$  separadas por una distancia  $h$  es cero (ecuación 2):

$$E[Z(x) - Z(x + h)] = 0 \quad [2]$$

y que las diferencias de varianza entre dos ubicaciones separadas por una distancia  $h$  dependen solamente del valor de  $h$  (ecuación 3).

$$E[\{Z(x) - Z(x + h)\}^2] = 2\gamma(h) \quad [3]$$

Donde  $\gamma(h)$  es la semivarianza de la variable  $Z$  a la distancia  $h$ . Estas dos condiciones, constancia de diferencias y varianza de diferencias (ecuaciones 2 y 3), definen lo que comúnmente se llama la hipótesis intrínseca de las variables regionalizadas ([Matheron 1962](#), [Matheron 1965](#)).

Por otra parte, el valor de la semivarianza puede ser estimado por medio de los datos de muestreo a través del cálculo del semivariograma experimental, cuya expresión viene dada por (ecuación 4):

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^N \{z(x_i) - z(x_i + h)\}^2 \quad [4]$$

donde,  $z(x_i)$  y  $z(x_i + h)$  son los valores del atributo  $Z$  en las localidades  $x_i$  e  $x_i + h$  y  $N$  el número de pares.

Es importante destacar que en el semivariograma experimental es posible observar las características espaciales del atributo en estudio, las que están dadas por un rango ( $A_0$ ) que representa la distancia  $h$ , en la cual la semivarianza tiende a ser constante y la dependencia espacial a ser nula, una cima ( $C$ ) que representa el valor máximo de semivarianza y un efecto pepita ( $C_0$ ) que corresponde al error o ruido aleatorio al interior de la distancia mínima entre puntos de muestreo. Estos tres valores son modelados matemáticamente por medio del ajuste de diferentes modelos que relacionan la distancia  $h$  con la semivarianza, siendo los más comunes: esférico, exponencial, gaussiano y lineal ([figura 1](#)).

Luego, en un proceso de interpolación espacial, la hipótesis básica de trabajo está basada en la observación común que, en promedio, valores de un atributo dentro de una vecindad tienen una fuerte probabilidad de ser similares y que esta probabilidad disminuye respecto a valores de otra vecindad separada por una gran distancia. Es decir, poseen las características de una variable regionalizada.

Este hecho es de alta importancia debido a que en un proceso de interpolación espacial el primer trabajo que se debe realizar es la comprobación de la hipótesis intrínseca de variable regionalizada, lo cual puede ser tratado por medio del análisis estadístico exploratorio, el que está basado en un

conjunto de herramientas estadísticas que permiten describir la estructura de los datos desde una perspectiva univariada, bivariada y espacial ([Kitanidis 1997](#)).

*Análisis estadístico exploratorio de datos:* En el análisis de una serie de datos, una de las primeras etapas que se lleva a cabo es la descripción de su distribución y cálculo de estadísticas básicas, de manera de verificar si ellos cumplen con las condiciones de base para la aplicación de un determinado modelo. Por ejemplo, en el caso de los modelos de productividad forestal, cuando la información ha sido tomada de un inventario, el primer análisis está orientado a verificar los supuestos asociados a una distribución normal, para luego proceder al cálculo de estadísticas descriptivas como media, varianza, desviación estándar, coeficiente de variación, cuantiles, mínimo y máximo entre otras. Posteriormente, es común ajustar un modelo y estimar valores para toda la población en estudio. Siempre bajo una hipótesis de partida que en esta situación es la normalidad de los datos.

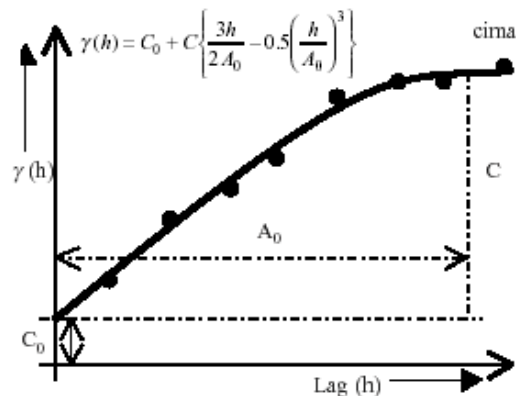
En el caso de un atributo espacial, para la aplicación de un modelo de interpolación, se parte de la base de que se está en presencia de una variable regionalizada, lo cual es comprobado a través de la aplicación de diferentes herramientas estadísticas a nivel univariado, bivariado y espacial ([Isaaks y Srivastava 1989](#)).

De acuerdo a [Kitanidis \(1997\)](#), un análisis exploratorio comienza con la obtención de un gráfico de los datos y posteriormente con el cálculo de sus estadísticas. Luego se deben estimar las características de la muestra de acuerdo a sus patrones de regularidad espacial.

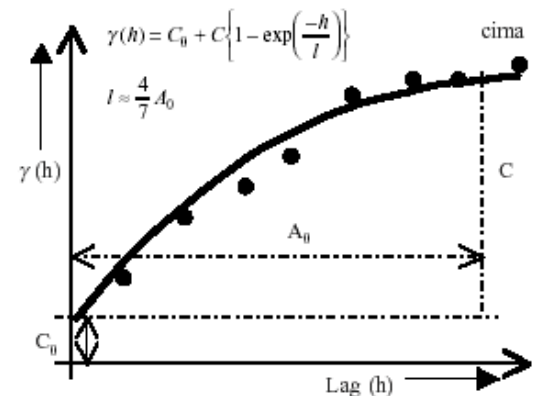
El análisis exploratorio comúnmente está dividido en dos partes: a) un análisis univariado y bivariado, usando la estadística tradicional, y b) un análisis de las relaciones espaciales utilizando geoestadística ([Atkinson 1999](#), [Goovaerts 1997](#), [Isaaks y Srivastava 1989](#), [Papritz y Stein 1999](#)).

El objetivo del análisis univariado y bivariado es describir y determinar el tipo de distribución de frecuencias e interrelaciones que poseen los atributos en estudio, siendo la distribución normal y lognormal las más comunes en fenómenos naturales. En este caso es común la realización de gráficas de frecuencias, cálculo de: media, varianza, covarianza, desviación estándar, rango y coeficientes estructurales. Además, es posible realizar diferentes pruebas no paramétricas como Kolmogorov-Smirnov, con la salvedad de que muchas de ellas asumen la independencia de las muestras ([Kitanidis 1997](#)).

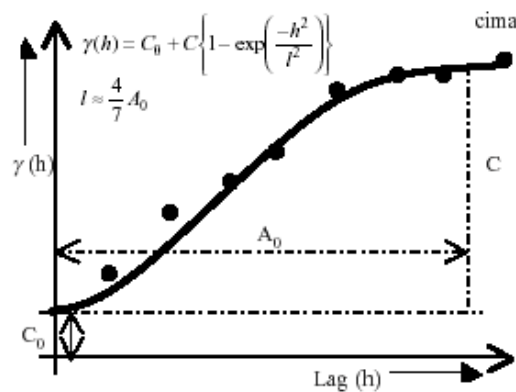
Por otra parte, el objetivo del análisis espacial es describir el comportamiento o estructura de un atributo sobre el espacio, para lo cual se determina el grado de dependencia espacial que presentan las observaciones en relación con sus vecinos y la continuidad espacial o rugosidad de la superficie. Los estimadores mayormente utilizados son las gráficas h-scatterplots, la covarianza, semivarianza y autocorrelación espacial. Una diferencia con respecto al análisis univariado y bivariado es que estos indicadores pueden ser ajustados de diferentes maneras sobre el espacio (isotrópico y anisotrópico), cambiando las distancias y ángulos sobre los cuales se hace el análisis ([Deutsch y Journel 1998](#)).



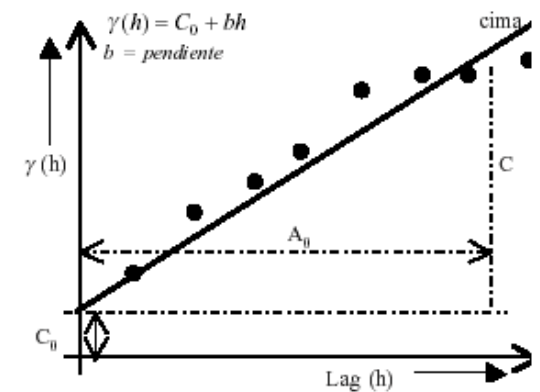
a) Modelo esférico



b) Modelo exponencial



c) Modelo gaussiano



d) Modelo lineal

Figura 1: Semivariograma experimental y modelos de semivarianza: a) esférico, b) exponencial, c) gaussiano y d) lineal.

Modeling an observed semi-variogram: a) spherical, b) exponential, c) gaussian, and d) linear.

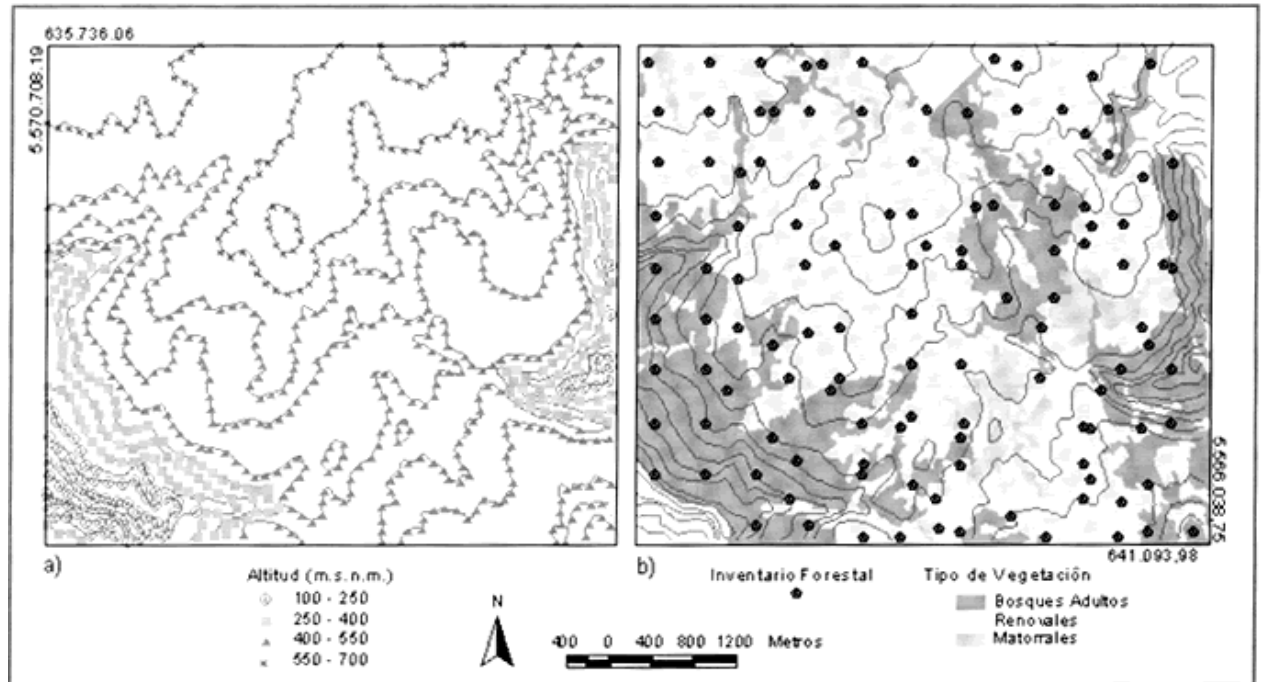
Finalmente, por medio del análisis exploratorio es posible detectar problemas de representatividad en las muestras, es decir, sectores con menor densidad de puntos respecto a otros, valores anómalos o existencia de grupos o subpoblaciones y tendencias en los datos. Todos estos aspectos están relacionados con el control antes de la aplicación de cualquier método de interpolación como los nombrados anteriormente.

## MATERIAL Y METODOS

**Area de estudio:** El estudio se ha localizado al interior de la Reserva Nacional Valdivia, ubicada a 20 km de la ciudad de Valdivia en la X Región de Los Lagos, Chile. Dicho sector tiene una superficie de 9.727 ha, presentando una topografía poco abrupta característica de la Cordillera de la Costa, con altitudes que van desde los 100 a 650 m s.n.m. La vegetación dominante corresponde principalmente a bosques naturales de los tipos forestales Siempreverde (subtipo Siempreverde y Renovales de Canelo) y Alerce, presentando diferentes estados de desarrollo y diversos niveles de intervención humana (CONAF *et al.* 1999, Donoso 1981, Rivera *et al.* 2002).

Para el estudio se seleccionó un área de 2.500 ha, de manera de obtener una situación representativa de la topografía del terreno y de la diversidad de formaciones vegetales presentes (figura 2).





*Figura 2:* Mapas de distribución de puntos de muestreo: a) altitud y b) área basal de árboles.  
Sample location maps: (a) altitude, and (b) basal area of trees.

Para ejemplificar la aplicación del análisis estadístico exploratorio en el proceso de interpolación se han escogido dos atributos con características espaciales y de intensidad de muestreo completamente opuestas. El primero corresponde a la altitud del terreno. Ella, comúnmente es representada en un modelo discreto basado en curvas de nivel, las que son empleadas para derivar un modelo continuo de la altitud utilizando interpolación espacial. Sobre las curvas de nivel, obtenidas a partir de cartas topográficas 1:50.000 del Instituto Geográfico Militar de Chile, se seleccionó una muestra compuesta de 816 puntos, los que fueron localizados a una equidistancia de 100 m sobre la trayectoria de cada curva, obteniendo una densidad promedio de 1 punto por cada 3 ha. ([figura 2a](#)).

El segundo corresponde al área basal (expresada en  $m^2/ha$ ) que fue obtenido de la red de inventarios forestales de la Reserva Nacional Valdivia ([Rivera et al. 2002](#)). Para el área de estudio se dispuso de 120 puntos con una densidad promedio de 1 punto cada 20 ha ([figura 2b](#)).

*Análisis univariado:* En una primera etapa, el análisis se orientó a describir el tipo de distribución de frecuencias y su grado de similitud con relación a una distribución normal que presentan los datos de altitud y área basal. Para esto se utilizó el análisis gráfico, medidas de la amplitud, de tendencia central, de variabilidad y de estructura. En forma anexa se realizó la prueba no paramétrica de Kolmogorov-Smirnov para comparar la forma de la distribución de muestras respecto a la forma de una función de densidad normal hipotética, dada una media y varianza ([cuadro 2](#)).

## CUADRO 2

Análisis univariado: lista de funciones estadísticas probadas.  
Univariate analysis: list of statistical functions tested.



Tipo de análisis	Medidas
Gráfico	Histograma y Función de Densidad
Amplitud	Mínimo, Máximo y Rango
Tendencia Central	Media y Mediana
Variabilidad	Desviación Estándar, Varianza y Coeficiente de Variación
Estructura	Grado de Skewness y Kurtosis
Prueba no paramétrica	Kolmogorov-Smirnov

*Análisis espacial:* En una segunda etapa, el análisis se orientó a describir la estructura y el grado de dependencia espacial que presentan los atributos altitud y área basal y a observar su variación en diferentes escalas, por lo cual se trabajó en dos niveles:

1. *Análisis regional de datos.* Se calculó el semivariograma experimental sobre toda la región de estudio (ecuación 3) y se ajustó un modelo de semivarianza ([cuadro 1](#)) donde se determinó:

- rango  $A_0$ , distancia a la cual la semivarianza es continua y la dependencia espacial entre los puntos tendería a ser mínima en un modelo isotrópico,
- rangos  $A_1$  y  $A_2$ , distancias donde la semivarianza es continua pero considerando los ejes mayor y menor en un modelo anisotrópico.
- cima  $C$ , valor máximo de semivarianza, y
- efecto pepita  $C_0$ , error aleatorio o ruido independiente. En forma paralela, se calculó el correlograma para determinar el grado de autocorrelación espacial de cada atributo en función de la distancia ( $h$ ) entre puntos ([Goovaerts, 1997](#)), (ecuación 5).

$$\rho(h) = \frac{C(h)}{\sqrt{\sigma_{+h}^2 * \sigma_{-h}^2}} \quad [5]$$

con:

$$C(h) = \frac{1}{N} \sum_{i=1}^N z(x_i) * z(x_i + h) - (m_{-h} * m_{+h})$$

$$m_{-h} = \frac{1}{N} \sum_{i=1}^N z(x_i)$$

$$m_{+h} = \frac{1}{N} \sum_{i=1}^N z(x_i + h)$$

$$\sigma_{-h}^2 = \frac{1}{N} \sum_{i=1}^N [z(x_i) - m_{-h}]^2$$

$$\sigma_{+h}^2 = \frac{1}{N} \sum_{i=1}^N [z(x_i + h) - m_{+h}]^2$$

Donde  $\rho(h)$  es la autocorrelación espacial a la distancia  $h$ , la que toma valores entre  $-1$  y  $1$ ,  $C(h)$  es la covarianza a la misma distancia y  $m_{-h}$ ,  $m_{+h}$ ,  $\sigma_{-h}^2$  y  $\sigma_{+h}^2$  son las medias y varianzas cola "tail" y cabeza "head" de la distribución ([Goovaerts 1997](#)).

El análisis consideró una distancia ( $h^{\max}$ ) de 5.409 m, con 10 intervalos (lag  $h$ ) de 540,9 m.

2. *Análisis local de datos.* Se calculó el semivariograma experimental (ecuación 3) y ajustó un modelo de semivarianza isotrópico para diferentes distancias de análisis  $h^{\max}$  (3.000 m, y 1.000 m) y se determinaron sus respectivos parámetros rango  $A_0$ , cima  $C$  y efecto pepita  $C_0$ .

Luego se calculó el correlograma de manera de determinar el grado de autocorrelación espacial existente a las mismas distancias  $h^{\max}$  (3.000 m y 1.000 m) (ecuación 5).

La selección de los modelos de semivarianza se realizó bajo tres criterios: minimizar la suma de los residuales (RSS), maximizar el coeficiente de correlación ( $R^2$ ) y verificando la existencia de isotropía o anisotropía en los datos.

## RESULTADOS Y DISCUSION

*Análisis univariado:* En el [cuadro 3](#) se presentan los valores de amplitud, tendencia central, variabilidad y estructura para los atributos altitud y área basal.

Respecto a la tendencia central de los datos, donde para una distribución normal la relación mediana/media alcanza el valor 1, se observa que para la altitud y área basal toma valores de 1,097 y 0,98 respectivamente. Esto muestra una diferencia de 9,7% y 2% frente a un valor ideal de normalidad, lo que por sí solo no es suficiente para decidir si los datos presentan o no una distribución normal. Este aspecto será retomado en la discusión.

Con relación a la variabilidad de los datos respecto a su tendencia central, se observa que la altitud presenta menos variabilidad que el área basal (coeficiente de variación de 26,7% y de 41,3% respectivamente). Este hecho puede ser explicado por las características naturales de los atributos o por problemas de representatividad de los datos.

Al nivel de la estructura, el área basal presenta un valor de skewness de 0,29. Una función normal de media 60,41 m<sup>2</sup>/ha y desviación estándar de 25,06 m<sup>2</sup>/ha presenta un valor de skewness de 0,22 ([cuadro 3](#)). Por otra parte, la altitud presenta un valor de -0,85, el que al ser negativo indica un desplazamiento hacia la derecha de las frecuencias respecto al valor medio de 455,76 m ([cuadro 3](#)).

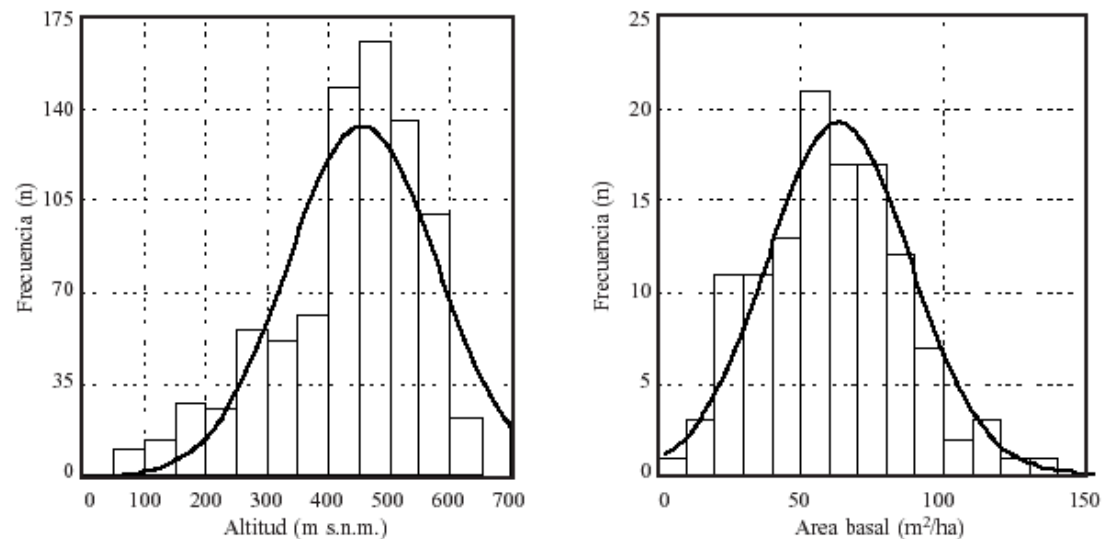
CUADRO 3

Caracterización estadística de los datos de altitud y área basal.  
Statistical characterization of altitude and data on the basal area of trees.

Atributo	Amplitud			Tendencia Central		Variabilidad			Estructura	
	Mínimo	Máximo	Rango	Media	Mediana	Varianza	Desviación Estándar	C.V (%)	Skewness	Kutosis
Altitud (m.s.n.m)	100	650	500	455,76	500	14.810,14	121,66	26,7	-0,85 (0,09)	0,23 (0,17)
Area Basal (m <sup>2</sup> /ha)	5,21	131,04	125,83	60,71	59,61	627,75	25,06	41,3	0,29 (0,22)	-0,14 (0,44)

En la [figura 3](#) se presenta el histograma y la función de densidad para los atributos altitud y área basal.

El análisis de esta figura corrobora los resultados presentados en el [cuadro 3](#), pero presenta también información complementaria. Por ejemplo, para la altitud se observa una diferencia entre la función de densidad y el histograma caracterizada por una subestimación de las frecuencias entre los intervalos [100, 300] m y [400, 600] m, además de una sobreestimación entre los intervalos [300, 400] m y [600, 650] m. En el caso del atributo área basal, sólo existe una subestimación puntual de la función de densidad entre los valores [20, 40] y [50, 60] m<sup>2</sup>/ha y una sobreestimación de la misma característica entre los valores [60, 70] y [100, 110] m<sup>2</sup>/ha. Finalmente, el bajo grado de Kurtosis presentado en el [cuadro 3](#) para la altitud y área basal respectivamente (0,23 y -0,14) es confirmado por la forma de los histogramas en las gráficas de la [figura 3](#) y por los valores cercanos a cero observados en la prueba de Kolmogorov- Smirnov (0,18 y 0,041 respectivamente para la altitud y área basal).



*Figura 3:* Histograma y función de densidad para atributos altitud y área basal.  
Histograms and density functions for altitude and the basal area of trees.

*Análisis espacial regional:* En la [figura 4](#) se presentan los variogramas y correlogramas correspondientes a los atributos altitud y área basal productos del análisis espacial en un nivel regional.

Al realizar un análisis espacial isotrópico de los atributos ([figura 4a](#)), se observa que la altitud posee un semivariograma caracterizado por un aumento paulatino de la semivarianza, lo que contrasta con un continuo y alto valor de semivarianza que presenta el atributo área basal en toda la región. De este comportamiento se puede deducir que en el caso del atributo altitud, valores vecinos separados a pequeñas distancias tenderían a ser semejantes, aumentando sus diferencias en función de la distancia entre ellos y en el caso del área basal, valores vecinos no serían semejantes debido a su alta variabilidad, la que se presenta independiente de la distancia, indicando la presencia de un alto valor de ruido aleatorio.

Estos comportamientos son corroborados al observar las gráficas de correlogramas ([figura 4b](#)), donde la altitud presenta valores altos de autocorrelación espacial en corta distancia (0,92 a 354 m), los que disminuyen hasta alcanzar valores cercanos a cero a una distancia de 2.100 m. En cambio, el área basal presentó siempre valores cercanos a cero independiente de la distancia.

Desde una perspectiva de análisis anisotrópico ([figura 4c y 4d](#)), se observa un comportamiento semejante al ya descrito para ambas variables, con la diferencia de que para el atributo altitud se acentúan los valores de semivarianza y autocorrelación espacial en los ejes de análisis 0° y 45°, conservando siempre la misma tendencia general del análisis isotrópico.

El ajuste de un modelo para describir la semivarianza de ambos atributos desde un análisis isotrópico e anisotrópico ([cuadro 4](#)) muestra para la altitud que una función esférica (isotrópica) explicaría en alto grado el comportamiento de la semivarianza ( $R^2 = 0,99$  y  $RSS = 2,67E+6$ ). De acuerdo a este modelo, la semivarianza aumenta paulatinamente tornándose estacionaria a una distancia ( $A_0$ ) de 4.430 m, con un valor máximo de  $19,550 \text{ m}^2$ . Se destaca que en este caso existe un bajo valor de ruido aleatorio o efecto pepita, el que sólo alcanza un 2,8% de la semivarianza total del modelo.

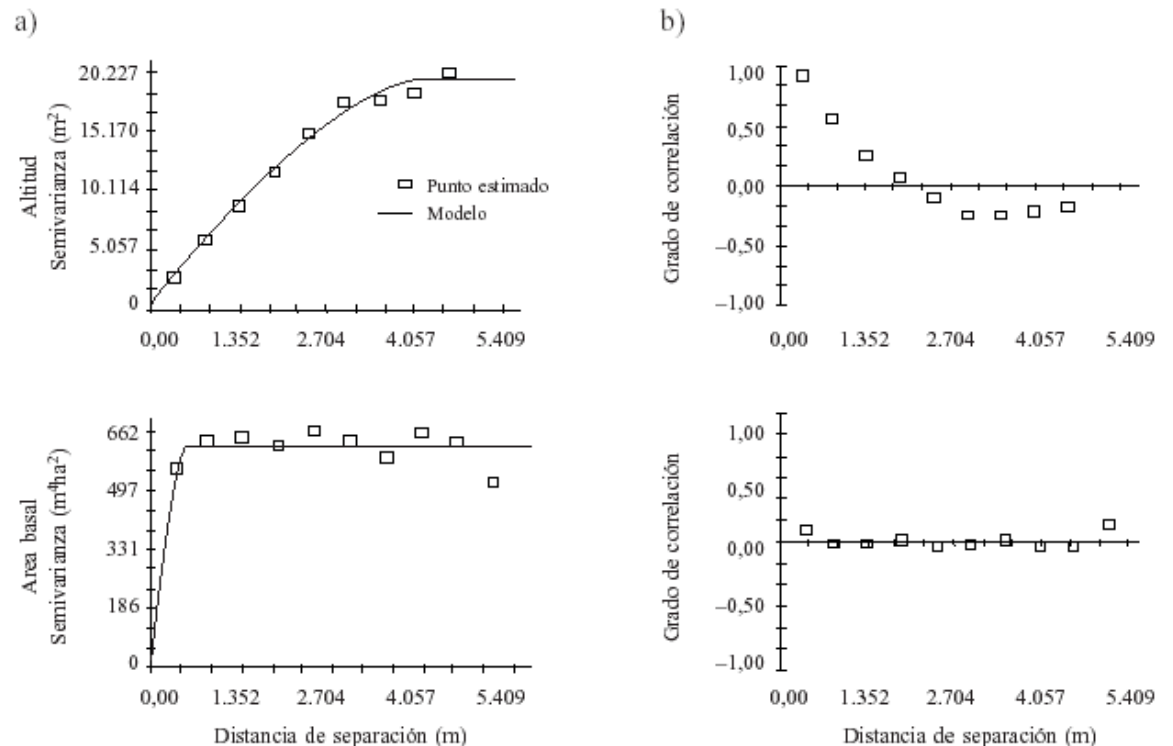
CUADRO 4

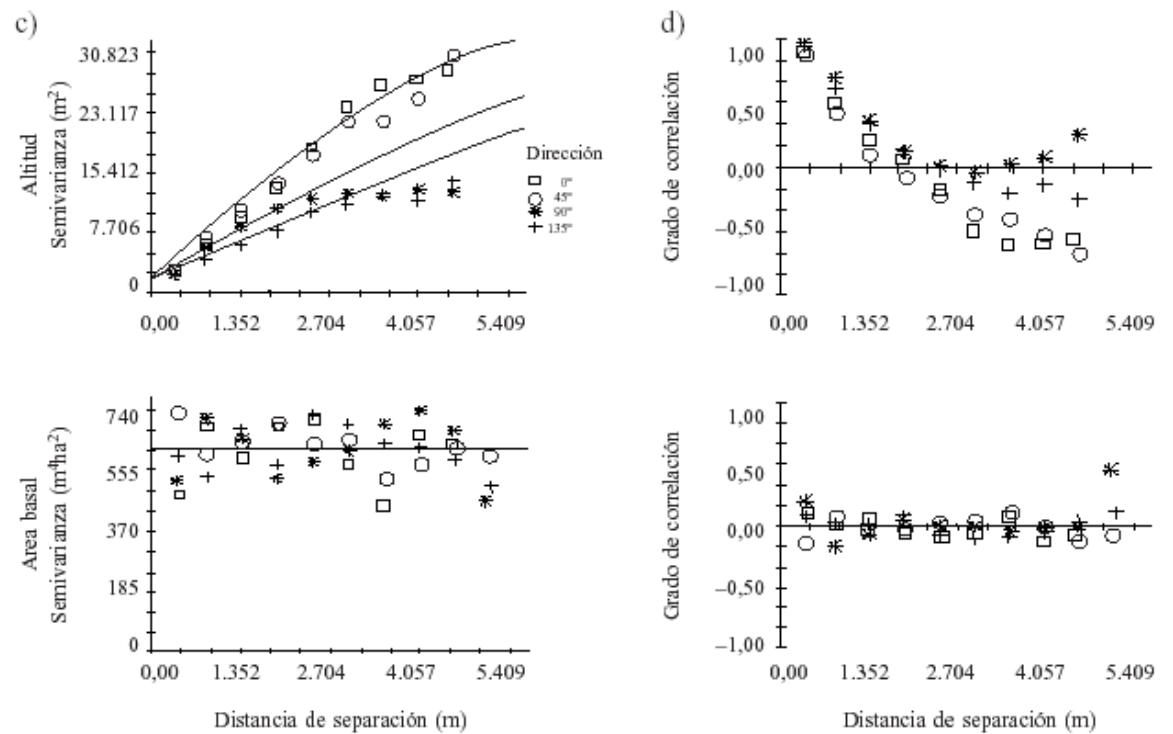
Análisis regional: modelos de semivarianza para la altitud y área basal.  
Regional analysis: variogram models for altitude and the basal area of trees.

Atributo	Análisis	Modelo	$C_0$	$C_0+C$	$A_0$	$A_1$	$A_2$	$R^2$	RSS
Altitud	Isotrópico	Esférico	550	19.550	4.430	–	–	0,99	2,67E+6
	anisotrópico	Esférico	1.770	32.973,15	–	12.820	6.310	0,93	4,72E+8
Area basal	Isotrópico	Esférico	20	622,1	560	–	–	0,21	14.541
	anisotrópico	Lineal	621,2	1.370,16	–	42.000	40.000	0,1	227.342

En el caso del análisis anisotrópico, el modelo de semivarianza muestra semejante comportamiento que el anterior, pero con mayores valores de ruido aleatorio ( $C_0$ ), de semivarianza ( $C$ ) y de rango ( $A_1$  y  $A_2$ ), produciendo básicamente una curva más extendida debido posiblemente a la presencia de una tendencia en las direcciones  $0^\circ$  y  $45^\circ$  por la diferencia en el número de pares (figura 4c).

Por otra parte, la semivarianza del área basal desde un análisis isotrópico fue modelada por medio de una función esférica, no lográndose un buen ajuste ( $R^2 = 0,21$  y  $RSS = 14.541$ ). Debido al bajo valor de rango ( $A_0 = 20$  m) que indica un aumento brusco de la semivarianza a una distancia muy por debajo de la distancia entre puntos de muestreo (447 m), se puede inferir que la semivarianza corresponde en su totalidad a ruido aleatorio ( $C_0$ ), siendo independiente de la distancia de análisis. Un comportamiento semejante presenta el análisis anisotrópico pero bajo un modelo lineal de la semivarianza (cuadro 4).

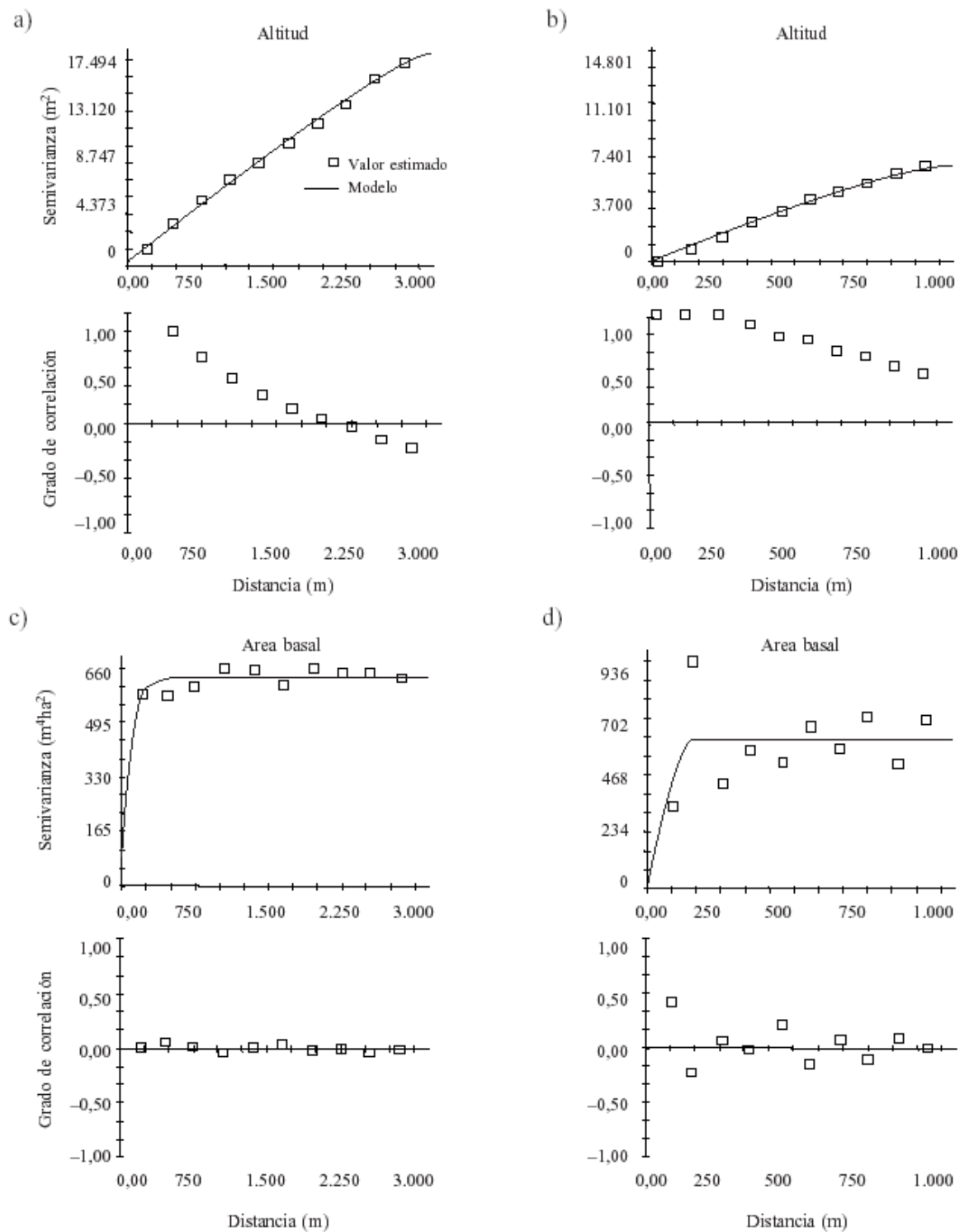




*Figura 4:* Análisis regional: Semivariogramas y correlogramas para la altitud y área basal: a) semivariogramas isotrópicos, b) correlogramas isotrópicos, c) semivariogramas anisotrópicos y d) correlogramas anisotrópicos.

Regional analysis: variograms and correlograms for altitude and the basal area of trees: (a) isotropic variogram, (b) isotropic correlogram, (c) anisotropic variogram, and (d) anisotropic correlogram.

En las [figuras 5](#) se presenta el variograma y correlograma de los atributos altitud y área basal para distancias locales de análisis  $h_{\max}$  de 3.000 m y 1.000 m.



*Figura 5:* Análisis local: Semivariogramas y correlogramas: a) altitud, distancia  $h_{max}$  de 3.000 m, b) altitud, distancia  $h_{max}$  de 1.000 m, c) área basal, distancia  $h_{max}$  de 3.000 m y d) área basal, distancia  $h_{max}$  de 1.000 m.

Local analysis: variograms and correlograms: (a) altitude, range up to 3000 m, (b) altitude, range up to 1000 m (c) tree diameters, range up to 3000 m, and (d) tree diameters, range up to 1000 m.

Para la altitud ([figura 5a y 5b](#)) se observa un comportamiento muy semejante al encontrado en un nivel regional de análisis, el que se caracteriza por un aumento paulatino de la semivarianza con la distancia, destacándose que no existe una mayor dispersión de sus valores debido al menor número de pares presentes en el análisis. Por otra parte, este comportamiento es corroborado por el ajuste de modelos de semivarianza ([cuadro 5](#)), donde en todos los casos se seleccionó una función esférica con valores semejantes de  $R^2$  y una pequeña disminución de los residuales. Se destaca que en el caso donde la distancia de análisis fue de 1.000 m existe una disminución de los valores de semivarianza total ( $C_0 + C = 7.130 \text{ m}^2$ ) y del rango ( $A_0 = 1.296 \text{ m}$ ), evidenciando la posibilidad de una subestructura (esférica) que operaría en distancias menores a 1.000 m ([figura 6](#)).

El área basal, por su parte, presenta siempre altos valores de semivarianza, los que se muestran independientes de la distancia empleadas en el análisis ([figura 5c y 5d](#)). Además, el ajuste de un modelo para la semivarianza lleva a valores bajos de  $R^2$  ([cuadro 5](#)), manteniéndose la tendencia de los valores encontrados en el análisis regional.

CUADRO 5

Análisis local: modelos de semivarianza para la altitud y área basal.  
Local analysis: variogram models for altitude and the basal area of trees.

Atributo	Distancia de análisis (m)	Modelo	$C_0$	$C_0 + C$	$A_0$	$R^2$	RSS
Altitud	5.409 (análisis regional)	Esférico	550	19.550	4.430	0,99	2,67E+6
	3.000	Esférico	330	21.760	4.923	0,98	825.477
	1.000	Esférico	10	7.130	1.296	0,99	335.827
Area basal	5.409 (análisis regional)	Esférico	20	622,1	560	0,21	14.541
	3.000	Esférico	77	632,3	94	0,27	7.033
	1.000	Esférico	1	615.8	167	0,28	189.045

Del análisis univariado podemos inferir que sólo la interpretación en conjunto de los datos de amplitud, tendencia central, variabilidad, estructura y gráficas de frecuencias permiten determinar que la altitud y área basal poseen una distribución muy cercana a la normal.

Por ejemplo, la diferencia entre los valores de media y mediana indican que ellas pertenecen a clases diferentes, pero no pudiéndose especificar en qué parte de la distribución se encuentran. Luego, el uso de un coeficiente de estructura como skewness permite conocer el sentido del desplazamiento de las frecuencias respecto al valor medio y gráficamente, la superposición de los histogramas con las funciones de frecuencias normales deja de manifiesto en forma explícita qué clases divergen de una distribución normal. Finalmente, estos comentarios son verificados de una forma cuantitativa al aplicar la prueba no paramétrica de Kolmogorov- Smirnov.

Por lo tanto, el uso del análisis univariado ha permitido determinar que la altitud y área basal son datos que respetan la primera parte de la hipótesis intrínseca de variables regionalizadas (ecuación 2). Para verificar si estos mismos datos respetan la segunda parte de ella (ecuación 3), es necesario efectuar un análisis espacial. Sólo con el cumplimiento de ambas es posible efectuar una interpolación espacial cuantitativamente aceptable.

En el caso de la altitud, la condición de –varianza de diferencias– se cumple, dado que el efecto pepita representa solamente un 2,8% de la semivarianza y tanto el rango como la cima están bien definidos ([figura 4a](#)). Este hecho, junto a la comprobación de la –constancia de diferencias–, evidentemente permite indicar que se respetan las condiciones básicas y es posible realizar una interpolación con este atributo ([Kitanidis 1997](#)).



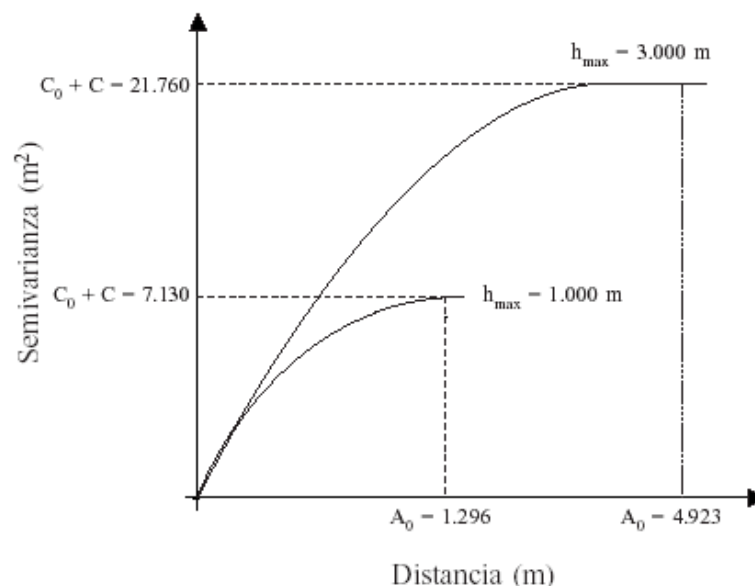
Ahora, el caso del área basal requiere una discusión más detallada. Debido a la falta de correlación entre los datos, mostrado por el correlograma correspondiente ([figura 4b](#)), no es posible ajustar de manera satisfactoria un modelo al variograma experimental. La única posibilidad sería un modelo con una sola componente (efecto pepita). Por ejemplo, la [figura 4a](#) presenta un modelo esférico donde los criterios de selección ( $R^2$  y RSS) no son adecuados. Por lo tanto, la hipótesis básica de interpolación –varianza de diferencias– no se respeta. No verificar esta situación antes de un proceso de interpolación podría llevar a resultados erróneos (independiente del algoritmo utilizado en el proceso).

Este hecho podría ser explicado, primero por un muestreo no apropiado a las características de variabilidad espacial del área basal. En otras palabras, las variaciones locales no son captadas dada la intensidad o distribución del muestreo sobre el espacio. [Gilbert y Lowell \(1997\)](#) encontraron resultados semejantes utilizando intensidades de muestreo de 16 puntos/ha y 4 puntos/ha en bosques boreales en Québec, Canadá. [Wallerman et al. \(2002\)](#) utilizan para un estudio exitoso de volumen de árboles una resolución espacial de muestras de 2,3 puntos/ha. pero agregando una estratificación de ellos basándose en fronteras naturales, permitiendo de esta manera un análisis más local.

Por otra parte, es posible que el atributo área basal no sea una variable regionalizada. En ese caso, la única posibilidad de producir planos de interpolación sería por medio de una triangulación de Delauney o diagramas de Voronoi ([Burrough y McDonnell, 1998](#)).

Dado lo anterior, es posible efectuar un análisis espacial a diferentes escalas sólo para altitud y no para el área basal.

Para la altitud, es posible observar la presencia de estructuras espaciales que estarían actuando a distancias de 0-1.000 m y de 1.000-5.409 m ([figura 6](#)). Como se observa en el [cuadro 5](#), los tres modelos poseen valores de  $R^2$  superiores a 0,9. Esta situación es de alta importancia ya que la semivarianza de la altitud podría ser caracterizada por un modelo esférico complejo que posee dos rangos y dos cimas, donde uno representaría la variabilidad a distancias menores que 1.000 m y otro la variabilidad a 5.000 m. [Currant y Atkinson \(1999\)](#) discuten sobre el efecto que tendrían las escalas y la resolución sobre las características espaciales de un atributo, lo cual es observado a través de modelos parciales de semivarianza.



*Figura 6:* Esquema de posible subestructura (modelo esférico) que se presenta en un nivel local para el atributo altitud.  
Suggested nested spherical model for altitude data.

Finalmente, el cumplimiento de las condiciones de –constancia de diferencias– y –varianza de diferencias – por parte de la altitud permiten decir que se está en presencia de una variable regionalizada. Esto asegura que dicho atributo pueda ser interpolado espacialmente dados los datos disponibles.

El resultado final de la interpolación dependerá principalmente de las características de los métodos empleados ([cuadro 1](#)), donde se debe considerar cómo es asumida la dependencia entre los puntos, vale decir, si ella tiene una forma lineal, cuadrática, potencia, inversa, entre otras.

## CONCLUSION

Con este estudio se ha mostrado que con el análisis estadístico exploratorio es posible describir la estructura espacial de los atributos altitud y área basal, lo que puede ser extensivo a otros del ámbito de recursos naturales. Además, ha permitido la verificación de los supuestos básicos para una interpolación (existencia de dependencia espacial) y, por otro lado, ha ayudado a la detección de problemas de representatividad de los datos. Finalmente, el análisis estadístico exploratorio se ha definido como un conjunto de herramientas del ámbito univariado y espacial, de fácil implementación e interpretación, que es la primera etapa para la obtención de buenos resultados en cualquier proceso de interpolación espacial.

## AGRADECIMIENTOS

Los autores agradecen al Sr. Hugo Rivera, Ingeniero Forestal, Coordinador del Plan de Ordenación de la Reserva Nacional Valdivia ([CONAFChile](#)), por haber facilitado los datos experimentales para el desarrollo de este estudio. Los autores agradecen también a los árbitros de este artículo, quienes con sus comentarios permitieron mejorar el texto. Finalmente, el primer autor agradece a la Pontificia Universidad Católica Chile y a CONICYT Chile por el soporte dado a sus estudios en Canadá.

## BIBLIOGRAFIA

ATKINSON, P. 1999. Spatial statistics. In: Stein, A., Van Der Meer, F. y Gorte, B. *Spatial Statistics for Remote Sensing*. Netherlands. Kluwer Academic Publishers. 264 p.

BURROUGH, P., R. MCDONNELL. 1998. *Principles of geographical information systems*. New York, Oxford University Press. 333 p.

CONAF, CONAMA, BIRF, UNIVERSIDAD AUSTRAL DE CHILE, PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE Y UNIVERSIDAD CATOLICA DE TEMUCO. 1999. Catastro y evaluación de los recursos vegetacionales nativos de Chile. Informe nacional con variables ambientales. Santiago, Chile. 88 pp.

DEUTSCH, C., A. JOURNAL. 1998. *Gslib: geostatistical software library and user's guide*. 2ª edición. New York. Oxford Press. 368 p.

CORVALAN, P., R. GOUET, C. REYES. 2000. Evaluación de estimaciones locales dasométricas realizadas mediante técnicas geoestadísticas. Seminario Silvotecná. Corporación Chilena de la Madera, Chile, 13 p.

CURRENT, P., P. ATKINSON. 1999. Issues of scale and optimal pixel size. En: Stein, A., Van Der Meer, F. y Gorte, B. *Spatial Statistics for Remote Sensing*. Netherlands. Kluwer Academic Publishers. 264 p.

DONOSO, C. 1981. Tipos forestales de los bosques nativos de Chile. En Investigación y desarrollo forestal. Conaf, Doc. Trabajo N° 30. Santiago, Chile. 70 pp.

- ISAAKS, E., M. SRIVASTAVA. 1989. *An introduction to applied geostatistics*. New York, Oxford University Press. 561 p.
- FELICISIMO, A. 1994. *Modelos digitales del terreno: principios y aplicaciones en las ciencias ambientales*. Madrid, Pentalfa Ediciones. 118 p.
- GARCIA, O. 1991. What is a diameter distribution? Iufro symposium on integrated forest management information systems. Tsukuba, Japan. 17 p.
- GILBERT, B., K. LOWELL. 1997. "Forest attributes and spatial autocorrelation and interpolation: effects of alternative sampling schemata in the boreal forest", *Landscape and Urban Planning*. 37: 235-244.
- GOOVAERTS, P. 1997. *Geostatistics for natural resources evaluation*. New York, Oxford University Press. 483 p.
- KITANIDIS, P. 1997. *Introduction to geostatistics: applications in hydrogeology*. New York, Cambridge University Press. 249 p.
- KYRIAKIDIS, P., J. DUNGAN. 2001. "A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions". *Environmental and Ecological Statistics*. 8: 311-330.
- LAM, N. 1983. "Spatial interpolation methods: a review". *The American Cartographer*. Vol. 10(2): 129-149.
- MATHERON, G. 1962. "Traité de géostatistique appliquée", *Memoires du Bureau de Recherches Géologiques et Minières* N°14. Paris. Éditions Technip. 333 p.
- . 1965. *Les variables régionalisées et leur estimation*. Paris, Masson et Cie. 305 p.
- OLIVER, M., 1990. "Kriging: a methode of interpolation for geographical information systems". *Int. J. Geographical Information Systems*. Vol. 4(3): 313-332.
- PAPRITZ, A., A. STEIN. 1999. Spatial prediction by linear kriging. In: Stein, A., Van Der Meer, F. y Gorte, B., 1999. *Spatial Statistics for Remote Sensing*. Netherlands. Kluwer Academic Publishers. 264 p.
- RIVERA, H., A. RUDLOFF, P. CRUZ. 2002. Plan de ordenación de la Reserva Nacional Valdivia. Una visión para el manejo ecológicamente sustentable de los ecosistemas forestales costeros de la X Región de Los Lagos. Proyecto manejo sustentable del bosque nativo. Santiago, Chile. Conaf/GTZ. 235 p.
- WALLERMAN, J., S. JOYCE, C. VENCATASAWMY, H. OLSSON. 2002. "Prediction of forest stem volume using kriging adapted to detected adges", *Canadian J. of Forest Research*. 32: 509-518.

Recibido: 25.06.02  
Aceptado: 03.03.03