

MODELOS PARAMÉTRICOS DE LA MATRIZ DE VARIANZAS-COVARIANZAS EN EL AJUSTE DE FUNCIONES DE PERFIL

J. G. Álvarez González, A. D. Ruiz González y M. Barrio Anta

Departamento de Ingeniería Agroforestal. Escuela Politécnica Superior. Universidad de Santiago de Compostela. Campus Universitario s/n. 27002-LUGO (España). Correo electrónico: algonjg@lugo.usc.es

Resumen

El empleo de funciones de perfil se ha generalizado durante las últimas décadas debido a su flexibilidad para estimar los volúmenes de árboles individuales hasta diferentes diámetros en punta delgada o con distintas longitudes de troza. Estas funciones se obtienen, generalmente, al ajustar modelos matemáticos a datos obtenidos por mediciones de diámetros a diferentes alturas en el tronco de un árbol. Se trata, por tanto de mediciones repetidas en cada una de las unidades de muestreo (árbol). La metodología comúnmente aplicada en el análisis de modelos de regresión con este tipo de datos se basa en establecer un modelo paramétrico para la matriz de varianzas-covarianzas. En este trabajo se analizan diversas estructuras de dicha matriz y se comparan empleando datos reales de análisis de troncos de *Pinus radiata* en Galicia.

Palabras clave: *Datos longitudinales, Dependencia espacial, Autocorrelación, Modelos espaciales*

INTRODUCCIÓN

Las funciones de perfil son ecuaciones que relacionan el diámetro o el radio del tronco de un árbol con la altura a la que se alcanza ese diámetro o radio. Su uso se ha generalizado durante las últimas décadas debido a su flexibilidad para estimar los volúmenes de árboles individuales hasta diferentes diámetros en punta delgada o con distintas longitudes de troza. Estas funciones se obtienen, generalmente, al ajustar modelos matemáticos a datos obtenidos por mediciones de diámetros a diferentes alturas en el tronco de un árbol, por lo que es razonable pensar que la variabilidad entre las medidas de cada árbol sea menor que entre árboles, no pudiéndose considerar por tanto como observaciones independientes. Es decir, existe una dependencia espacial (autocorrelación) en los datos y los errores, con lo que se

viola un supuesto básico para estimar un modelo de regresión mediante mínimos cuadrados ordinarios, que asume que los errores son independientes e idénticamente distribuidos.

Para solventar el problema de la autocorrelación en el ajuste han surgido diferentes metodologías, siendo los modelos con coeficientes aleatorios y los modelos paramétricos de la matriz de varianzas-covarianzas de los datos las más utilizadas (GREGOIRE *et al.*, 1995; TASSISA & BURKHART, 1998; GARBER & MAGUIRE, 2003). En este trabajo se ha empleado esta última metodología porque presenta algunas ventajas importantes para el ajuste de funciones de perfil (NUÑEZ-ANTÓN Y ZIMMERMAN, 2001): (i) permite obtener de una forma más eficiente los estimadores de los parámetros del modelo que relaciona la variable dependiente con las variables independientes; (ii) permite obtener estimadores

más adecuados para los errores estándar de los estimadores de dichos parámetros; (iii) en muchos casos, permite solucionar de forma efectiva los problemas de falta de datos o de diferente número de mediciones entre individuos; y (iv) puede emplearse aún cuando el número de mediciones sea muy grande en comparación con el número de individuos.

MATERIAL Y MÉTODOS

Los datos empleados en este trabajo provienen de un total de 30 árboles seleccionados aleatoriamente entre la muestra de 365 árboles con que se ajustó la función de perfil de la especie en Galicia (CASTEDO Y ÁLVAREZ, 2000). La muestra original proviene de 61 parcelas establecidas en masas regulares de *Pinus radiata* con edades comprendidas entre 5 y 41 años. Dichas parcelas se han distribuido por toda Galicia tratando de cubrir las diferentes calidades de estación y de modo que, en la muestra final de árboles medidos, estuviesen adecuadamente representadas las principales combinaciones de diámetros normales y alturas. Cada árbol fue apeado y dividido en trozas de longitud comprendida entre 1 y 2,5 metros, siendo en su mayoría de un metro de longitud.

En cada troza se midió la longitud con cinta métrica y aproximación al centímetro, el diámetro basal en cruz con forcípula milimétrica y el espesor de corteza radial en milímetros. Así mismo se midieron el diámetro normal y el diámetro a 4 metros en cruz, ambos en milímetros, la altura total con cinta métrica centimétrica y, por último, se estimaron los volúmenes totales con y sin corteza cubicando cada troza mediante la fórmula de Smalian.

Por tanto, los datos procedentes del análisis de tronco corresponden a mediciones de diámetro en alturas no equidistantes entre sí ni coincidentes de unos individuos a otros, es decir, se trata de datos no rectangulares, lo que impone una limitación a algunas de las matrices de varianzas-covarianzas a utilizar en la modelización de la estructura del error. En la figura 1, donde se muestran los perfiles de los 30 árboles seleccionados, se puede observar como las alturas de medición no son las mismas en todos los individuos y las distancias entre unas mediciones y otras tampoco son constantes.

Para evitar este inconveniente se ha empleado la función de perfil ajustada a la especie en Galicia para estimar los valores de los diámetros relativos (cociente entre el diámetro a una cierta altura relativa y el diámetro normal) a 10 alturas

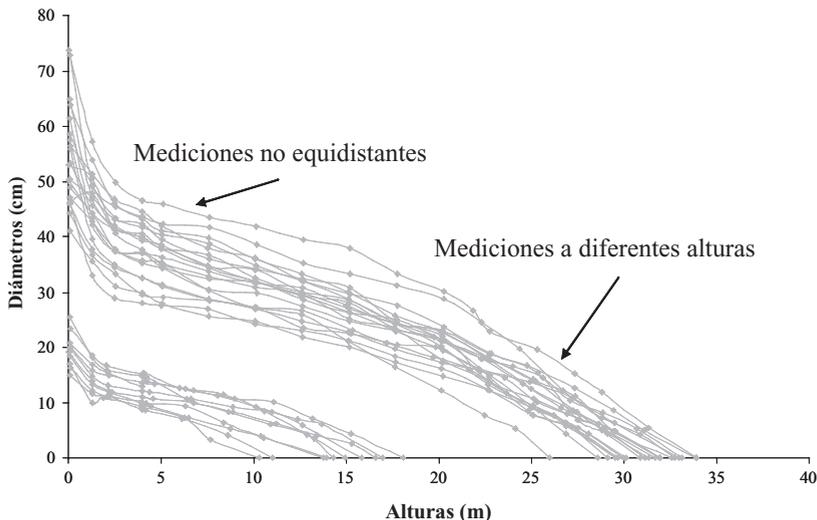


Figura 1. Perfiles de los 30 árboles seleccionados en los que se observa que las mediciones no se realizan a las mismas alturas y las distancias entre mediciones no son constantes

relativas (cociente entre la altura del árbol a la que se mide el diámetro y la altura total) equidistantes e iguales para cada árbol correspondientes a dividir la altura total en diez partes iguales. Por tanto las 10 alturas relativas empleadas para cada árbol se corresponden con valores que van desde 0,1 hasta 1 con una equidistancia de 0,1. La expresión matemática de la función de perfil ajustada a la especie es la siguiente (CASTEDO Y ÁLVAREZ, 2000):

$$r = \frac{b_0 \cdot D}{1 - e^{-b_2 \cdot (1,3-H)}} + \left(\frac{D}{2} - b_0 \cdot D \right) \cdot \left[1 - \frac{1}{1 - e^{-b_1 \cdot (1,3-H)}} \right] + (1) \\ + e^{-b_1 \cdot h} \cdot \left[\frac{\left(\frac{D}{2} - b_0 \cdot D \right) \cdot e^{1,3 \cdot b_1}}{1 - e^{-b_1 \cdot (1,3-H)}} \right] - e^{-b_2 \cdot h} \cdot \left[\frac{b_0 \cdot D \cdot e^{-b_2 H}}{1 - e^{-b_2 \cdot (1,3-H)}} \right]$$

donde *r* es el radio con corteza correspondiente a la altura *h*, en cm; *D* es el diámetro normal con corteza, en cm; *h* es la altura en metros desde la base del árbol hasta el punto donde se alcanza el diámetro *d*; *H* es la altura total del árbol, en *m* y *b*₀, *b*₁ y *b*₂ son los parámetros obtenidos en el ajuste y cuyos valores son 0,45034; 0,76913 y 0,05076, respectivamente.

En la tabla 1 se muestran los datos de las medias y las varianzas de los diámetros relativos a las diferentes alturas relativas obtenidos al emplear la función de perfil. También se muestran los valores de los coeficientes de correlación

entre los valores de los diámetros relativos a las diferentes alturas relativas. Se puede observar como el valor medio de los diámetros relativos decrece, como es lógico, por lo que no se trata de un valor estacionario y, por tanto, es necesario emplear un modelo para representar dicha variación. Por otra parte, la varianza de los diámetros relativos muestra variaciones aunque, con el propósito de simplificar el análisis se va a considerar estacionaria, es decir, se asume que existe homogeneidad de varianzas. En cuanto a los valores del coeficiente de correlación lineal, se observa que existen fuertes correlaciones entre todos los diámetros relativos, que los valores decrecen muy ligeramente a medida que las mediciones están más alejadas y que las correlaciones sub-diagonales se pueden considerar estacionarias, es decir, las correlaciones entre cualquier diámetro relativo y el diámetro relativo a *n* mediciones anteriores (donde *n* varía de 1 a 9) es constante.

Finalmente se cuenta con dos bases de datos para realizar los ajustes: la original, con mediciones no equidistantes y con diferente número para cada individuo; y la construida con la función de perfil en la que las mediciones son equidistantes en altura relativa y hay el mismo número por individuo.

Entre las matrices empleadas con la base de datos original para corregir los posibles problemas de dependencia de los errores se han incluido modelos espaciales que tienen en cuenta la

Altura relativa	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
Medias	0,93	0,84	0,77	0,69	0,61	0,52	0,41	0,29	0,16
Varianzas:	0,0002	0,0002	0,0001	0,0002	0,0001	0,0002	0,0001	0,0001	0,0001
N	30	30	30	30	30	30	30	30	30
Corr.:	1,0000								
	0,9998	1,0000							
	0,9996	0,9999	1,0000						
	0,9993	0,9998	0,9999	1,0000					
	0,9989	0,9996	0,9998	0,9999	1,0000				
	0,9984	0,9993	0,9996	0,9998	0,9999	1,0000			
	0,9979	0,9989	0,9994	0,9996	0,9998	0,9999	1,0000		
	0,9973	0,9984	0,9990	0,9994	0,9996	0,9998	0,9999	1,0000	
	0,9965	0,9979	0,9985	0,9990	0,9993	0,9996	0,9998	0,9999	1,0000

Tabla 1. Valores medios y varianzas de los diámetros relativos de los 30 árboles a alturas relativas equidistantes. Se presentan también las correlaciones entre dichos diámetros relativos y la línea discontinua muestra como las correlaciones sub-diagonales se pueden considerar constantes

Matrices empleadas para la base de datos original	
Modelo simple sin dependencia de errores	Modelo potencial
$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 \end{bmatrix}$
Modelo exponencial	Modelo Gaussiano
$\sigma^2 \begin{bmatrix} 1 & e^{-d_{12}/\theta} & e^{-d_{13}/\theta} & e^{-d_{14}/\theta} \\ e^{-d_{21}/\theta} & 1 & e^{-d_{23}/\theta} & e^{-d_{24}/\theta} \\ e^{-d_{31}/\theta} & e^{-d_{32}/\theta} & 1 & e^{-d_{34}/\theta} \\ e^{-d_{41}/\theta} & e^{-d_{42}/\theta} & e^{-d_{43}/\theta} & 1 \end{bmatrix}$	$\sigma^2 \begin{bmatrix} 1 & e^{-d_{12}^2/\rho^2} & e^{-d_{13}^2/\rho^2} & e^{-d_{14}^2/\rho^2} \\ e^{-d_{21}^2/\rho^2} & 1 & e^{-d_{23}^2/\rho^2} & e^{-d_{24}^2/\rho^2} \\ e^{-d_{31}^2/\rho^2} & e^{-d_{32}^2/\rho^2} & 1 & e^{-d_{34}^2/\rho^2} \\ e^{-d_{41}^2/\rho^2} & e^{-d_{42}^2/\rho^2} & e^{-d_{43}^2/\rho^2} & 1 \end{bmatrix}$
Modelo lineal	
$\sigma^2 \begin{bmatrix} 1 & 1-\rho \cdot d_{12} & 1-\rho \cdot d_{13} & 1-\rho \cdot d_{14} \\ 1-\rho \cdot d_{21} & 1 & 1-\rho \cdot d_{23} & 1-\rho \cdot d_{24} \\ 1-\rho \cdot d_{31} & 1-\rho \cdot d_{32} & 1 & 1-\rho \cdot d_{34} \\ 1-\rho \cdot d_{41} & 1-\rho \cdot d_{42} & 1-\rho \cdot d_{43} & 1 \end{bmatrix} \rho \cdot d_{ij} \leq 1$	
Matrices empleadas para la base de datos construida a partir de la función de perfil	
Modelo simple sin dependencia de errores	Modelo de simetría compuesta
$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$
Modelo autorregresivo de orden 1	Modelo autorregresivo de media móvil
$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$	$\sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ \gamma & 1 & \gamma & \gamma\rho \\ \gamma\rho & \gamma & 1 & \gamma \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{bmatrix}$
Modelo Toeplitz en una banda	Modelo Toeplitz en dos bandas
$\sigma^2 \begin{bmatrix} 1 & \sigma_1 & 0 & 0 \\ \sigma_1 & 1 & \sigma_1 & 0 \\ 0 & \sigma_1 & 1 & \sigma_1 \\ 0 & 0 & \sigma_1 & 1 \end{bmatrix}$	$\sigma^2 \begin{bmatrix} 1 & \sigma_1 & \sigma_2 & 0 \\ \sigma_1 & 1 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & 1 & \sigma_1 \\ 0 & \sigma_2 & \sigma_1 & 1 \end{bmatrix}$

Tabla 2. Matrices empleadas para modelizar la estructura del error en la base de datos original y en la base de datos construida a partir de la función de perfil. Donde σ^2 es la varianza, d_{ij} es la distancia entre la medición i y la medición j y los restantes parámetros (σ_1 , σ_2 , ρ y γ) definen el tipo de dependencia entre los errores

influencia que tiene la distancia entre las mediciones en la correlación entre dichos errores. Estas matrices no han sido utilizadas en el caso de la base de datos originada con la función de perfil. En la tabla 2 se muestran las expresiones de las matrices utilizadas en el caso de que sólo se contase con cuatro mediciones.

Puesto que el objetivo de este estudio no es analizar los resultados de diferentes funciones de perfil sino caracterizar la estructura del error, para modelizar la estructura de la media se ha empleado un modelo totalmente saturado, es decir, un modelo lineal en el que intervienen todas las variables que afectan al proceso, que en este caso son la altura en la base de datos original y la altura relativa en la base de datos construida a partir de la función de perfil.

Para realizar los ajustes se han asumido una serie hipótesis de partida: (i) las mediciones realizadas en árboles distintos son independientes entre sí; (ii) la variable dependiente está normalmente distribuida y (iii) las varianzas de los errores son homogéneas. La estimación de los parámetros se ha realizado por máxima verosimilitud restringida o residual (para más detalles ver LITTLE et al., 1996) empleando el procedimiento MIXED del programa SAS (SAS INSTITUTE INC., 2001). Por último, la comparación entre modelos se ha basado en el cálculo de los valores del Cri-

terio de Información de Akaike (AICd) y del Criterio de Información Bayesiano (BICd) en diferencias (BURNHAM & ANDERSON, 2000).

RESULTADOS Y DISCUSIÓN

En la tabla 3 se muestran los resultados obtenidos para el ajuste de los datos originales, es decir, con distinto número de observaciones por individuo y con mediciones no equidistantes entre sí. Como era de esperar, los peores resultados corresponden al modelo simple de matriz de varianzas-covarianzas en el que no se tiene en cuenta la dependencia de errores. Con el resto de los modelos la precisión del ajuste es bastante similar, aunque los valores más bajos de los criterios de información de Akaike y Bayesiano en diferencias se obtienen al emplear un modelo potencial para la matriz de varianzas-covarianzas. En este modelo se asume que la dependencia entre las mediciones decrece a medida que estas se van espaciando, y la relación entre ellas viene dada por un término de correlación (ρ) elevado a la distancia, en valor absoluto, que separa ambas mediciones en el tronco del árbol.

En la tabla 4 se muestran los resultados obtenidos al emplear la base de datos construi-

Modelo	AICd	BICd
Simple	12,0	13,4
Potencial	0,0	0,0
Exponencial	2,1	3,4
Gaussiano	2,1	3,4
Lineal	3,3	3,9

Tabla 3. Resultados del ajuste de diferentes matrices de varianzas-covarianzas a la base de datos original

Modelo	AICd	BICd
Simple	547,6	544,8
Simetría compuesta	0,0	0,0
Autorregresivo orden 1	14,0	12,6
Autorregresivo de media móvil	280,5	279,1
Toeplitz de una banda	471,9	469,1
Toeplitz de dos bandas	305,6	304,2

Tabla 4. Resultados del ajuste de diferentes matrices de varianzas-covarianzas a la base de datos construida a partir de la función de perfil

da a partir de la función de perfil. De nuevo, los peores resultados corresponden al modelo simple de matriz varianzas-covarianzas, poniendo de relieve que la inclusión de un factor que tenga en cuenta la dependencia entre mediciones mejora la precisión del modelo ajustado. Los mejores resultados se han obtenido con el modelo de simetría compuesta en el que se asume que la correlación entre los errores es constante. Esto puede deberse a la fuerte estacionariedad que se apreciaba en la base de datos originada mediante el empleo de la función de perfil, pero que en realidad no es frecuente en este tipo de estudios, por lo que no parece aconsejable emplear este tipo de estructura a no ser que un análisis previo de los datos indique lo contrario. Por otra parte, el modelo autorregresivo de orden 1 también presenta buenos resultados con la ventaja de que las correlaciones no son constantes, sino que su valor decrece a medida que las mediciones se van alejando a lo largo del tronco.

De estos resultados se puede concluir que: (i) independientemente de la estructura de datos empleada, la modelización del error mejora la precisión del modelo; (ii) cuando se emplea una estructura de datos original parece conveniente modelizar la estructura del error empleando modelos espaciales sencillos, como por ejemplo un modelo potencial; y (iii) en el caso de que se cuente con datos rectangulares y equidistantes en el espacio, un modelo autorregresivo de orden 1 o un modelo de covarianzas constantes podría ser suficiente, dependiendo de la estacionariedad de las correlaciones entre mediciones.

Por último, destacar que todos los modelos utilizados pueden adaptarse de forma sencilla para tener en cuenta una posible falta de homogeneidad de varianzas en los datos de partida, aunque esto podría suponer mayores dificultades para lograr la convergencia del método de estimación de los parámetros.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Ciencia y Tecnología mediante el proyecto “Crecimiento y evolución de las masas de pinar en Galicia” del Plan Nacional I+D+I (2000-2003).

BIBLIOGRAFÍA

- BURNHAM, K.P. & ANDERSON D.R.; 2000. *Model selection and inference: a practical information-theoretic approach*. Springer-Verlag. Berlin.
- CASTEDO, F. Y ÁLVAREZ GONZÁLEZ, J.G.; 2000. Construcción de una tarifa de cubicación con clasificación de productos para *Pinus radiata* D. Don en Galicia basada en una función de perfil del tronco. *Inv. Agr., Sist. Rec. For.* 9(2): 253-268.
- GARBER, S.M. & MAGUIRE D.A.; 2003. Modelling stem taper of three central Oregon species using mixed effects models and autoregressive error structures. *For. Ecol. Manage.* 179: 507-522.
- GREGOIRE, T.G.; SCHABENBERGER, O. & BARRET, J.P.; 1995. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent plot measurements. *Can. J. For. Res.* 25: 137-156.
- LITTELL, R.C.; MILLIKEN, G.A.; STROUP, W.W. & WOLFINGER R.D.; 1996. *SAS System for Mixed Models*. SAS Institute Inc. Cary, NC.
- NUÑEZ-ANTÓN, V. Y ZIMMERMAN, D.L.; 2001. Modelización de datos longitudinales con estructuras de covarianza no estacionarias: modelos de coeficientes aleatorios frente a modelos alternativos. *Qüestió* 25(2): 225-262.
- SAS INSTITUTE INC.; 2001. *SAS/STAT™ User's Guide, Release 8.2*. Cary, N.C. USA.
- TASSISA, G. & BURKHART H.E.; 1998. An application of mixed effects analysis to modeling thinning effects on stem profile of loblolly pine. *For. Ecol. Manage.* 103: 87-101.