

# **FUNCIONES DE CORRELACIÓN ESPACIAL PARA MODELOS MIXTOS ADITIVOS DE ÁRBOL INDIVIDUAL**

## **SPATIAL CORRELATION FUNCTIONS FOR INDIVIDUAL TREE MIXED MODELS WITH ADDITIVE GENETIC EFFECTS**

**Eduardo Pablo Cappa<sup>1</sup>**  
**Rodolfo J. C. Cantet<sup>2</sup>**

<sup>1</sup>Ing. Agr. Facultad de Agronomía, UBA. Estudiante doctoral, becario FONCyT.

E-mail: ecappa@mail.agro.uba.ar

<sup>2</sup>Ing. Agr. M.S., Ph. D. Facultad de Agronomía, UBA. E-mail: rcantet@mail.agro.uba.ar

### **SUMMARY**

Modeling environmental (co)variability is a relevant issue in genetic evaluation of forest trees: observations that are closer in space are more intensely correlated than more distant ones. We review different spatial correlation functions that are useful for predicting breeding values of trees by means of mixed models.

**Key words:** mixed models, spatial analysis, correlation function.

### **RESUMEN**

Uno de los problemas más complejos en la evaluación genética forestal consiste en la modelación de la (co)variabilidad ambiental: observaciones de árboles más cercanos en el espacio se encuentran más intensamente correlacionadas que observaciones más distantes. En esta presentación se revisan distintas funciones de correlación espacial apropiadas para el análisis de ensayos genéticos forestales mediante modelos mixtos.

**Palabras clave:** modelos mixtos, análisis espacial, función de correlación.

### **INTRODUCCIÓN**

A diferencia de lo que ocurre con cultivos anuales o con animales, la evaluación genética forestal muestra una extensa heterogeneidad ambiental a nivel de micro y macrositio (Costa e Silva *et al*, 2001). El único modo de evitar, o de aliviar, el efecto de la (co)variabilidad ambiental sobre la predicción del valor de cría de los árboles es emplear una modelación apropiada. El análisis espacial es una metodología estadística apta para considerar la correlación espacial

En este trabajo consideraremos distintas *funciones de correlación* potencialmente apropiadas para predecir el valor de cría de árboles mediante modelos mixtos (Henderson, 1984). A tal efecto emplearemos el enfoque del “análisis de vecino más cercano” (ej. Cullis y Gleeson, 1991). En la sección 2 presentaremos un modelo aditivo mixto de árbol individual con errores correlacionados espacialmente. En la sección 3, examinaremos distintas funciones de correlación espacial. Finalmente, en la sección 4 se presenta una discusión general.

## MODELO LINEAL MIXTO ADITIVO DE ÁRBOL INDIVIDUAL CON ERRORES CORRELACIONADOS ESPACIALMENTE

La expresión matricial de la ecuación del modelo aditivo mixto de árbol individual utilizado en la evaluación genética forestal para predecir el valor de cría es igual a:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{w} + \mathbf{e} \quad [1]$$

donde  $\mathbf{y}$  es el vector de observaciones para la característica bajo análisis;  $\mathbf{b}$  es el vector de efectos fijos sistemáticos, asociado a los datos mediante la matriz de incidencia  $\mathbf{X}$ ; el vector aleatorio  $\mathbf{a}$  contiene los valores de cría o efectos genéticos aditivos y posee distribución  $N_q(\mathbf{0}, \mathbf{A}s^2_A)$ , donde  $\mathbf{A}$  es la matriz de relaciones aditivas entre árboles (Henderson, 1984) y  $s^2_A$  la varianza genética aditiva. La asociación entre  $\mathbf{a}$  y los datos se realiza mediante la matriz de incidencia  $\mathbf{Z}$ . Una característica distintiva del modelo [1] es la inclusión del vector de variables aleatorios  $\mathbf{w}$  correlacionados espacialmente, que siguen una distribución  $N_n(\mathbf{0}, \mathbf{H}s^2_w)$ . En la sección siguiente describiremos distintas estructuras posibles que puede tomar la matriz  $\mathbf{H}$ . Finalmente,  $\mathbf{e}$  es el vector de errores aleatorio desconocido con distribución  $N_n(\mathbf{0}, \mathbf{I}s^2_e)$  que incluyen efectos genéticos no aditivos, errores de medición y errores de efecto de micrositio (considerado este a la distancia menor a la espaciamiento entre árboles).

### ESTRUCTURAS DE LA MATRIZ $\mathbf{H}$

Para relacionar los elementos de la matriz  $\mathbf{H}$  ( $H_{ij}$ ) con la *distancia entre árboles* ( $d_{ij}$ ), utilizaremos una *función de correlación* ( $\rho(d_{ij})$ ) tal que:

$$H_{ij} = 1 - r(d_{ij})$$

donde  $d_{ij}$  es la distancia, un número real no-negativo y  $\rho(d_{ij})$  debe ser una *función de correlación* positiva definida (Abrahamsen, 1997). Adicionalmente,  $\rho(d_{ij})$  posee las siguientes características (Diggle, 2000):

1. *Monótona decreciente*: la correlación entre los elementos de  $\mathbf{w}$  disminuye a medida que la distancia entre ellos aumenta.
2.  $r(d_{ij}) \rightarrow 0$  cuando  $d_{ij} \rightarrow \infty$ : la correlación tiende a cero cuanto más separado esté el par de variables en  $\mathbf{w}$ ,
3. al menos un parámetro controla la disminución de la correlación con la distancia.

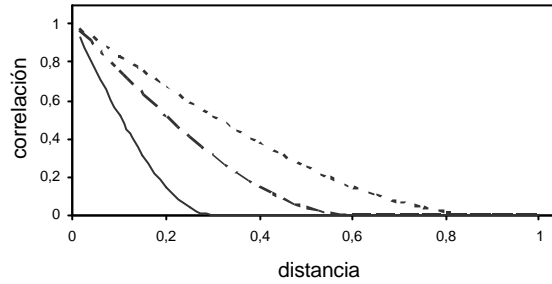
Estas especificaciones son válidas cuando la variabilidad es independiente de la *dirección* y sólo dependen de la distancia  $d$  (Zimmerman and Harville, 1991). La función  $\rho(d_{ij})$  incluye uno o dos parámetros que le proveen cierta flexibilidad para captar el decaimiento en covarianza con la distancia. A continuación examinaremos algunas de estas funciones.

### Función esférica

Esta *función de correlación* es igual a:

$$r(d_{ij}; ?) = \begin{cases} 1 - \frac{3}{2} \frac{d_{ij}}{q_s} - \frac{1}{2} \left( \frac{d_{ij}}{q_s} \right)^3 & \text{para } 0 \leq d_{ij} \leq q \\ 0 & \text{para } q > d_{ij} \end{cases}$$

El parámetro  $?_s$  modifica la caída en la correlación, como se observa en la Figura 1. Note que, a medida que  $?_s$  aumenta (de 0.3 a 0.6 y de 0.6 a 0.9), la correlación cae más suavemente con el aumento en la distancia, hasta hacerse casi lineal. Sin embargo, la dependencia de un sólo parámetro de escala ( $?_s$ ), le da escasa flexibilidad a la disminución de la correlación.



**Figura 1. Función de correlación esférica con  $\phi_s = 0.3$  (línea llena),  $\phi_s = 0.6$  (línea cortada) y  $\phi_s = 0.9$  (línea punteada).**

La palabra ‘esférica’ en el nombre se refiere a que  $\rho(d_{ij})$  se interpreta, geoméricamente, como la intersección de dos esferas en tres dimensiones cuyos centros se encuentran a una distancia igual a  $d$ . Fu *et al.* (1999) utilizaron la *función de correlación* esférica para modelar patrones de variación de la altura en ensayos de progenies de *Pseudotsuga menziesii* (Mirb.) Franco. Asimismo, Hamann *et al.* (2002) encontraron que la función esférica para los efectos espaciales, permitió aumentar la precisión de los estimadores de los efectos de procedencia en *Alnus rubra* Bong. Como ejemplo de una  $\rho(d_{ij})$  esférica, y para  $\phi_s = 0.6$ , presentamos la siguiente matriz de correlación:

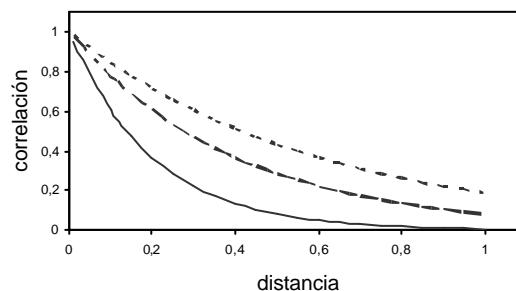
1	0.518	0.148	0	0	0
0.518	1	0.518	0.148	0	0
0.148	0.518	1	0.518	0.148	0
0	0.148	0.518	1	0.5185	0.148
0	0	0.148	0.518	1	0.518
0	0	0	0.148	0.518	1

### Función exponencial

Esta familia de *funciones de correlación* esta definida por la expresión:

$$r(d_{ij}) = \exp\left(-\frac{d_{ij}}{q_x}\right)$$

siendo  $q_x > 0$ . Como en el caso de la función esférica, el parámetro  $q_x$  modifica también la caída en la correlación. Sin embargo, como se observa en la Figura 2, dicha disminución es más suave para un mismo valor del parámetro. Por ejemplo, cuando  $q_s = q_x = 0.6$ , los elementos  $H_{12}$  y  $H_{13}$  con  $\rho(d_{ij})$  esférica, son iguales a 0.518 y 0.148. Mientras que, si  $\rho(d_{ij})$  es exponencial, dichas correlaciones son 0.717 y 0.513. Cuando  $i = 1$  y  $j = 6$ , el elemento  $H_{ij}$  es 0 y 0.189 para  $\rho(d_{ij})$  esférica y exponencial, respectivamente.



**Figura 2. Función de correlación exponencial con  $\phi_x = 0.2$  (línea llena),  $\phi_x = 0.4$  (línea cortada) y  $\phi_x = 0.6$  (línea punteada).**

La matriz de correlación de la  $\rho(d_{ij})$  exponencial cuando  $\rho_x = 0.6$  es igual a:

$$\begin{bmatrix} 1 & 0.717 & 0.513 & 0.368 & 0.264 & 0.189 \\ 0.717 & 1 & 0.717 & 0.513 & 0.368 & 0.264 \\ 0.513 & 0.717 & 1 & 0.717 & 0.513 & 0.368 \\ 0.368 & 0.513 & 0.717 & 1 & 0.717 & 0.513 \\ 0.264 & 0.368 & 0.513 & 0.717 & 1 & 0.717 \\ 0.189 & 0.264 & 0.368 & 0.513 & 0.717 & 1 \end{bmatrix}$$

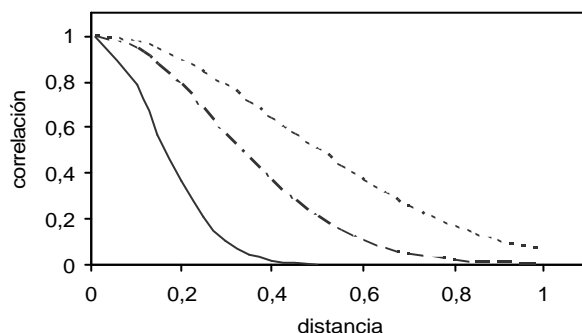
La estructura es equivalente a un proceso autorregresivo de primer orden (Cullis y Gleeson, 1991). Costa e Silva *et al.* (2001) y Dutkowski *et al.* (2002) emplearon la  $\rho(d_{ij})$  exponencial en modelos aditivos de árbol individual.

### Función gaussiana

Una alternativa a la *función de correlación* exponencial es la gaussiana cuya expresión es igual a:

$$r(d_{ij}) = \exp\left(-\left(\frac{d_{ij}}{q_g}\right)^2\right)$$

siendo  $q_g > 0$ . Note que la única diferencia entre la  $\rho(d_{ij})$  exponencial y la gaussiana es el cuadrado en el exponente. Como se observa en la Figura 3 la  $\rho(d_{ij})$  gaussiana muestra una caída de correlación muy suavizada cercano a  $d = 0$ , presentando un alto grado de continuidad en el origen: es infinitamente diferenciable. Las diferencias entre los elementos contiguos de  $H_{ij}$  ejemplifican mejor el impacto de elevar al cuadrado dentro del término exponencial, cuando se trata de valores cercanos a  $d = 0$ . Tomando nuevamente  $q_x = q_g = 0.6$ , la diferencia  $H_{11} - H_{12}$  para la  $\rho(d_{ij})$  exponencial es 0.284. Alternativamente, cuando  $\rho(d_{ij})$  es gaussiana,  $H_{11} - H_{12} = 0.105$ . A distancias mayores la correlación exponencial es más alta. Por ejemplo,  $H_{16}$  es 0.189 para  $\rho(d_{ij})$  exponencial y 0.062 para la gaussiana.



**Figura 3. Función de correlación gaussiana con  $q_g = 0.2$  (línea llena),  $q_g = 0.4$  (línea cortada) y  $q_g = 0.6$  (línea punteada).**

La matriz de correlación gaussiana para  $q_g = 0.6$  es la siguiente:

$$\begin{bmatrix} 1 & 0.895 & 0.641 & 0.368 & 0.169 & 0.062 \\ 0.895 & 1 & 0.895 & 0.641 & 0.368 & 0.169 \\ 0.641 & 0.895 & 1 & 0.895 & 0.641 & 0.368 \\ 0.368 & 0.641 & 0.895 & 1 & 0.895 & 0.641 \\ 0.169 & 0.368 & 0.641 & 0.895 & 1 & 0.895 \\ 0.062 & 0.169 & 0.368 & 0.641 & 0.895 & 1 \end{bmatrix}$$

La  $\rho(d_{ij})$  gaussiana ha sido escasamente utilizada para modelar *funciones de correlación* en el espacio. Quizás esto sea debido a la dificultad de encontrar procesos biológicos espaciales donde la correlación disminuye tan lentamente cerca de  $d = 0$ .

## DISCUSIÓN

La inclusión de una *función de correlación* espacial en un modelo mixto aditivo de árbol individual produjo mejoras en la exactitud de predicción de los valores de cría (Costa e Silva *et al.*, 2001; Dutkowski *et al.*, 2002). En esta presentación se han revisado tres  $\rho(d_{ij})$  para modelar la variación espacial a escala local, en el análisis de ensayos genéticos forestales. El cálculo de las predicciones BLUP (Henderson, 1984) de los valores de cría requieren estimar los parámetros de dispersión ( $s^2_A$ ,  $s^2_w$ ,  $s^2_e$  y el parámetro de suavizado  $q$  para la  $\rho(d_{ij})$  considerada), sea por métodos de máxima verosimilitud o empleando teoría bayesiana con cadenas de Markov por simulación de Monte Carlo (MCMC). Investigaciones futuras se concentrarán en dilucidar cuales de estas  $\rho(d_{ij})$  mejoran la exactitud de predicción de los valores de cría, siendo el cálculo de las ecuaciones de modelo mixto y la estimación paramétrica computacionalmente factible, en especial para el parámetro  $q$ .

## BIBLIOGRAFIA

- ABRAHAMSEN, P. (1997) "A Review of Gaussian random fields and correlation functions". Second edition. Rapport Nr. 917. ISBN 82-539-0435-5.
- COSTA e SILVA, J., Dutkowski G. W., and Gilmour A. R.. (2001). "Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual." *Canadian Forest Research* 31(11): 1887-1893.
- CULLIS, B.R., and Gleeson A. C. (1991). "Spatial Analysis of Field Experiments – An Extension to two Dimensions" *Biometrics* 47: 1449-1460.
- DIGGLE, P. J. and Ribeiro JR, P.J. (2000) Model-based Geostatistics. Associação Brasileira de Estatística. 137pp.
- DUTKOWSKI, G. W., Costa e Silva J., Gilmour A.R, and Lopez G. A. (2002). "Spatial analysis methods for forest genetic trials." *Canadian Forest Research* 32: 2201-2214.
- FU, Yong-Bi, Yanchuk A. D., and Namkoong G. (1999). "Spatial patterns of tree height variations in a series of Douglas-fir progeny trials: implications for genetic testing." *Canadian Forest Research* 29: 714-723.
- HAMANN, A., Namkoong G. and Koshy M.P. (2002). "Improving Precision of Breeding Values by Removing Spatial Autocorrelated Variation in Forestry Field Experiments." *Silvae Genetica* 51(5-6): 210-215.
- HENDERSON, C. R. (1984). Applications of Linear Models in Animal Breeding. Canadá, University of Guelph, ON. 462pp.
- ZIMMERMAN D.L., Harville D.A. (1991). "A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments" *Biometrics* 47: 223-239.