

ANÁLISIS ESPACIAL DE ENSAYO DE FAMILIAS DE POLINIZACIÓN ABIERTA DE *EUCALYPTUS TERETICORNIS* SMITH

López, C.¹(*); Salto, C.²; Venturini, M.³; Ewens, M.⁴

¹Facultad de Ciencias Forestales/ Universidad Nacional de Santiago del Estero.
Av. Belgrano (S) 1912, (4200) Santiago del Estero – Argentina.

²Instituto Nacional de Tecnología Agropecuaria, E.E.A. Concordia, Entre Ríos

³Facultad de Ciencias Forestales/ Universidad Nacional de Santiago del Estero

⁴Estación Forestal Fernández/ Universidad Católica de Santiago del Estero

Autor para la correspondencia: carlos@unse.edu.ar

Boletín del CIDEU 6-7: 47-55 (2008)
ISSN 1885-5237

Resumen

La modelación de la variación espacial del diámetro normal en un ensayo de 51 familias de polinización abierta de *Eucalyptus tereticornis* SMITH, con un proceso separable autorregresivo de primer orden de residuales, mostró ser más eficiente que el modelo básico de bloques completos al azar (BCA) para la predicción de valores genéticos en la aplicación de modelos lineales mixtos. Fueron utilizados variogramas de residuales como herramienta de diagnóstico de la variación ambiental alineada en filas y columnas del ensayo. La inclusión de los efectos fijos y aleatorios en los modelos fue testada mediante la prueba de F y la de razón de verosimilitud (LRT) respectivamente. Los componentes de varianza fueron estimados por máxima verosimilitud restringida (REML) y la solución de los valores de mejora predichos usando el mejor predictor lineal insesgado (BLUP). El análisis espacial promovió un mejor ordenamiento y selección de los genotipos.

Palabras clave: *Bloques completos aleatorizados, proceso autorregresivo, REML, BLUP, modelo mixto.*

Summary

Spatial análisis in an open-pollinated family trial of *Eucalyptus tereticornis* Smith

Modeling spatial variation of diameter in a 51 open-pollinated family trial of *Eucalyptus tereticornis* SMITH, with a separable first order autoregressive process (AR1), showed more efficient than the randomized complete blocks design to predict genetics values by a linear mixed model.

Variograms of residuals were used as a diagnosis tool of environmental variation aligned in columns and rows in the trial.

Fixed and random effects in the model were tested by means of the F test and the likelihood ratio test (LRT) respectively. Variance components were estimated by restricted maximum likelihood (REML) and breeding values predicted using best linear unbiased prediction (BLUP). Spatial analysis showed to be more suitable to rank and select genotypes.

Key words: *Randomized complete blocks, first order autoregressive process, REML, BLUP.*

I. Introducción

El objetivo básico del mejoramiento genético es acumular combinaciones de genes que favorezcan los rasgos de interés a través de la selección (Ipinza, 1998). La selección explota la variabilidad existente procurando favorecer la reproducción de individuos con características deseables (Falconer, 1981). La evaluación de la calidad genética de estos individuos se realiza a través de pruebas de progenies que tienen, principalmente, la finalidad de predecir sus valores de mejora sobre la base del desempeño de su descendencia (Gezán y Torres, 1998).

Para incrementar la precisión en la comparación de los tratamientos genéticos es necesario lograr condiciones ambientales homogéneas con el diseño experimental. Para este cometido es común efectuar el control local mediante el establecimiento de bloques completos aleatorizados (Mora Garcés y Ramírez, 2000). Sin embargo, el elevado número de tratamientos que normalmente incluyen los ensayos genéticos forestales conspira contra ese supuesto debido a que el aumento del tamaño de los bloques disminuye su capacidad para controlar la variación espacial de las condiciones del sitio. Esto conduce a la reducción de la capacidad de los análisis para identificar variaciones genéticas y la precisión en las estimaciones de valores genéticos (Costa e Silva *et al.*, 2001).

La variabilidad del sitio en los ensayos de campo puede ser espacialmente continua, reflejando patrones similares en el suelo subyacente y efectos microclimáticos; discontinua, reflejando efectos de cultivo y de medición; o aleatoria debida a la heterogeneidad microambiental (Dutkowski *et al.*, 2002). La asociación espacial positiva (gradiente ambiental) o negativa (competencia) entre las observaciones

(Apiolaza, 2006) induce a una posible asociación entre unidades experimentales vecinas que contradice el supuesto de independencia de los registros.

Se sugirieron varios enfoques analíticos que toman en cuenta la variación del sitio y mejoran la estimación de los efectos de los tratamientos en los ensayos de campo (Dutkowski *et al.*, 2002). Entre los antecedentes de estos métodos se cuenta el modelo propuesto por Papadakis en 1937 que efectúa una corrección por una covarianza que utiliza los residuos o los rendimientos de las unidades experimentales (Mora Garcés y Ramírez, 2000). Los análisis que asumen la existencia de una estructura de covarianza residual próxima a la estructura de covarianza real de los datos conducirán a mejorar la eficiencia y precisión en la estimación de los efectos de los tratamientos (Costa e Silva *et al.*, 2001). Gilmour *et al.*, (1997) propone un modelo autorregresivo de primer orden de residuales que asume un proceso separable en filas y columnas. En este contexto, este trabajo propone los siguientes objetivos:

Objetivo General

Examinar la contribución del análisis espacial en la evaluación de experimentos genéticos.

Objetivos Específicos

- Comparar el modelo BCA con modelos espaciales
- Modelar los residuos espacialmente correlacionados
- Seleccionar el modelo apropiado
- Predecir y estimar los promedios familiares y ganancias esperadas

II. Materiales y métodos

El material consta de un ensayo de progenies de 51 familias de polinización

Tabla 1. Descripción del material genético.

Procedencia	Grado de Mejora	Identificación	Origen	Latitud	Longitud	Altitud
Anhembí (San Pablo)	APS-F1	T3 B75	QLD: Mareeba, Helenvale, Mt. Garnete Cooktown	22° 8' S	48° 11' 67"	457 m
Anhembí (San Pablo)	APS-F2	T3 E89	QLD: Cooktown, Helenvale, Mareeba, Mt. Garnet	22° 8' S	48° 11' 67"	457 m

abierta de *Eucalyptus tereticornis* procedentes de San Pablo, Brasil (Tabla 1).

Área de Estudio

El área de estudio es la Estación Experimental Fernández, Santiago del Estero, Argentina, ubicada a 63° 55' Longitud Oeste; 27° 55' Latitud Sur y 160 m.s.n.m.

Diseño Experimental

El ensayo fue establecido en 1996 utilizando el diseño experimental de bloques completos aleatorizados con 4 repeticiones y parcelas lineales de 5 árboles. El distanciamiento inicial fue de 3 m x 2 m y la bordura perimetral simple. Los datos corresponden a árboles individuales localizados en las celdas de una cuadrícula formada por filas y columnas. La forma regular del ensayo se consiguió con el agregado de celdas con valores perdidos.

Evaluación del ensayo

Se evaluó el diámetro normal con corteza medido con cinta en el décimo año.

Análisis Estadístico

Los análisis se realizaron con modelos lineales mixtos que responden a la forma general:

$$y = X\beta + Zu + e$$

En esta expresión, y es el vector de los datos de cada árbol, β y u son los vectores de los efectos fijos y aleatorios respectivamente; e es el vector de los efectos residuales aleatorios; y X y Z son las matrices de incidencia que relacionan las observaciones a los efectos del modelo. Se asumió que la distribución conjunta de los efectos aleatorios tiene distribución normal multivariada con promedio y (co)varianza:

$$\begin{bmatrix} u \\ e \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right)$$

donde 0 es una matriz nula y G y R son matrices de varianza-covarianza correspondientes a u y e , respectivamente.

En este marco se utilizó un modelo genético familiar de acuerdo con la siguiente expresión lineal (Borralho, 1998):

$$y_{ij} = \mu + f_i + e_{ij}$$

donde: y_{ij} representa las observaciones registradas de diámetro, μ es el promedio general; f_i el efecto genético aditivo aleatorio, y e_{ij} el efecto residual.

El modelo familiar incluye los valores de mejora de las madres en la matriz u e incorpora la matriz de parentesco en el

análisis permitiendo la estimación directa de la varianza genética aditiva y los valores de mejora.

Las soluciones de los efectos fijos y aleatorios se obtuvieron resolviendo las ecuaciones de modelos mixtos (Henderson, 1984):

$$\begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \end{bmatrix}$$

donde \mathbf{R} es la matriz de varianza-covarianza de los residuos y \mathbf{G} es la suma directa de las matrices de varianza-covarianza de cada uno de los efectos aleatorios.

Si los residuos se asumen independientes, \mathbf{R} se define como $\sigma_e^2 \mathbf{I}$. Sin embargo, los análisis espaciales permiten que \mathbf{R} tenga diferentes estructuras de correlación basada en la descomposición de e :

$$e = \xi + \eta$$

donde ξ y η se refieren a los residuos espacialmente correlacionados y residuos aleatorios independientes, respectivamente.

La estructura de ξ puede ser modelada de varias maneras. Para datos ordenados en filas y columnas, Gilmour *et al.*, (1997) sugieren modelarla como un proceso autorregresivo de primer orden (AR1 x AR1) con una matriz \mathbf{R} :

$$\mathbf{R} = \sigma_\xi^2 [\text{AR1}(\rho_c) \otimes \text{AR1}(\rho_r)] + \sigma_\eta^2 \mathbf{I}$$

donde σ_ξ^2 es la varianza residual espacial; AR1 son las matrices de correlación autorregresivas de primer orden con parámetros de autocorrelación ρ_c y ρ_r para columnas y filas respectivamente y \otimes es el producto Kronecker. La matriz de varianza-covarianza residual σ_η^2 es la varianza de los residuos independientes e \mathbf{I} es la matriz

identidad. Los elementos del vector η se asumen independientes y reflejan los efectos ambientales debido a las diferencias entre micrositos, efectos genéticos no aditivos y errores de medición (Costa Silva *et al.*, 2001).

Fueron utilizados varias combinaciones de efectos aleatorios en la aplicación de los modelos. La inclusión de los efectos aleatorios fue testada mediante la prueba de razón de verosimilitud (LRT) (Searle, 1971) que prueba el doble de la diferencia del logaritmo de la verosimilitud con un valor tabulado de Chi cuadrado de dos colas al 95 % de probabilidades. También, fue extendido el modelo espacial para incluir términos que controlen la variación alineada en filas y columnas.

Los componentes de la varianza fueron estimados por máxima verosimilitud restringida (REML), los valores de mejora predichos usando el mejor predictor lineal insesgado (BLUP) y el rendimiento promedio de las familias fueron estimados con mejor estimador lineal insesgado (BLUE).

III. Resultados y discusión

El análisis exploratorio y genético de los datos promueve la siguiente discusión de los resultados:

Fue comprobado el supuesto de normalidad de los datos mediante la prueba de Shapiro - Wilks (prob = 0,30).

Modelos de predicción

En la Tabla 2 se muestran los ajustes de los modelos de predicción BCA y espaciales probados. En los modelos básicos de BCA los cambios del logaritmo de la verosimilitud debido a la inclusión de bloques y parcelas son inferiores al valor crítico de 3,84 ($p = 0,05$) lo que denota diferencias no significativas de acuerdo a la prueba de LRT (Likelihood Ratio Test). Por

Tabla 2. Logaritmos de máxima verosimilitud (REML log-likelihood) de los modelos de predicción

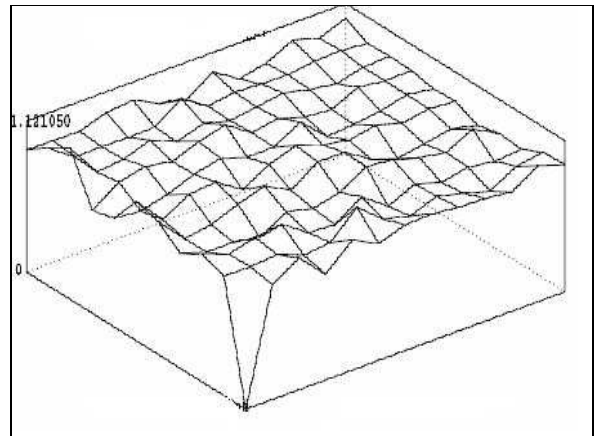
Efecto aleatorio del modelo	Modelo del residuo	REML Log-likelihood		Varianza del residuo	Error estándar de la diferencia
			Cambio		
Familia + bloque + parcela	BCA	-73,619	--	15,098	1,493
Familia + bloque	BCA	-73,861	--	15,612	1,488
<i>Familia</i>	BCA	-73,861	0,24	15,612	1,488
Familia	AR1 x AR1	-73,221	0,64	15,663	1,495
Familia + columna	AR1 x AR1	-70,161	3,7	14,851	1,487

lo tanto es preferible el modelo básico que no incluye bloques ni parcelas.

La aplicación del modelo espacial (AR1 x AR1) tampoco muestra un cambio significativo (0,64) respecto del modelo básico adecuado. Sin embargo, la inclusión de las columnas como factor aleatorio que controla esa variación espacial es altamente significativa, como muestra el cambio de 3,7 en el Logaritmo de la Verosimilitud. Esta diferencia produce un LRT de 7,40, significativamente superior al valor crítico, lo que determina su selección para modelar la variación ambiental.

Si bien no es una prueba concluyente, el menor valor que acusa el error estándar de la diferencia (1,487) sugiere un mejor ajuste del modelo. Este estadístico no es una prueba formal de ajuste porque es muy influenciado por el modelo espacial particular. Asimismo, el variograma de los residuales que muestra la Figura 1 corrobora el control de la variación espacial. El variograma, que sirve como herramienta de diagnóstico para evaluar la presencia de variación extraña (Gilmour *et al.*, 1997), no acusa patrones en los residuos.

Las predicciones de los promedios familiares son diferentes en los modelos básicos y espaciales. La comparación de las

**Figura 1. Variograma de residuales**

primeras 20 familias ordenadas por su desempeño predicho se muestran en la Figura 2. La magnitud de los promedios predichos y el ordenamiento de los tratamientos, tiene un efecto directo sobre la selección de los genotipos en los programas de mejoramiento genético.

Los parámetros y las ganancias estimadas se expresan en la Tabla 3. La estimación de la heredabilidad familiar (h^2) en sentido estricto es considerablemente alta, reflejando alto control genético del diámetro normal. Rocha *et al.*, (2006) estimaron valores de heredabilidad que varían desde 0.14 hasta 0.64 en diámetro normal en *Eucalyptus urophylla*. Un valor de 0.45 fue obtenido por Valera y

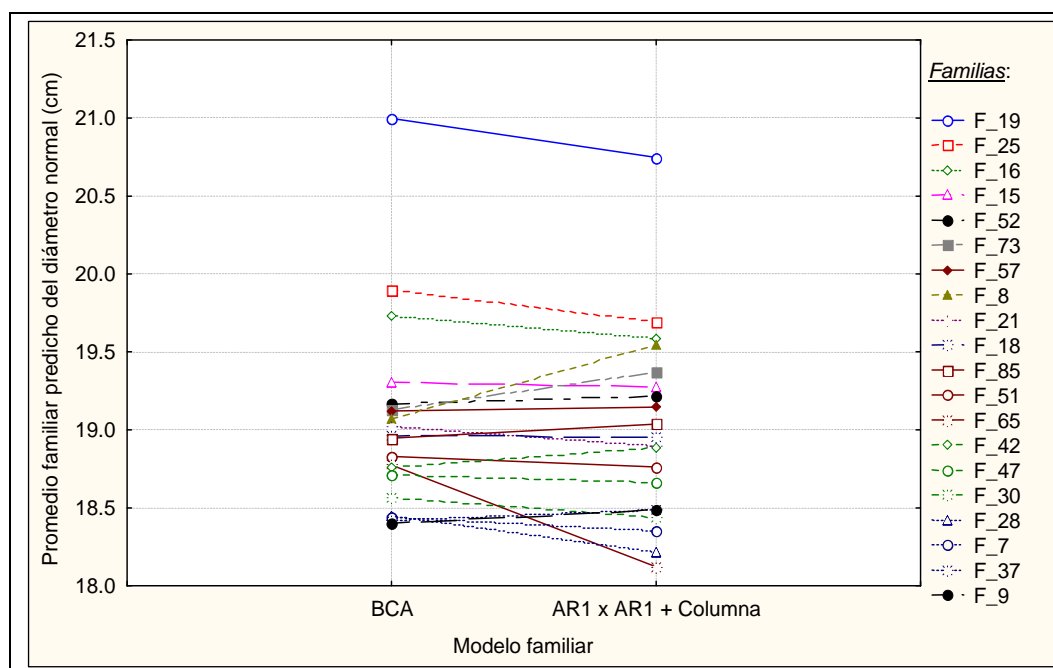


Figura 2. Comparación de promedios familiares predichos por los modelos

Tabla 3. Promedio y parámetros genéticos. Los errores estándar se indican entre paréntesis.

Modelo	Parámetros			
	Promedio	h^2	Varianza aditiva	Ganancia (%)
BCA	18,03 (0,3)	0.568 (0.09)	2.30 (0.8)	5,57
Espacial	18,03 (0,3)	0,459 (0.11)	2.24 (0.8)	5,48

Tabla 4. Logaritmos de máxima verosimilitud de los modelos de estimación

Efecto aleatorio del modelo	Modelo del residuo	REML Log-likelihood		Varianza del residuo	Error estándar de la diferencia
			Cambio		
bloque	BCA	-808,36	--	15,43	1,112
Fila	AR1 x AR1	-807,58	0,78	15,56	1,06
Columna	AR1 x AR1	-804,65	3,71	14,641	1,05

Kageyama, (1995) en *Eucalyptus saligna* a los 32 meses de edad.

La varianza genética aditiva constituye un indicativo de la facilidad de identificación de genotipos superiores que proporcionarán ganancias con la selección. Para facilitar su valorización se usa el coeficiente de

variación genética aditiva (CVa) que es una ponderación de la varianza genética respecto al promedio del rasgo. El coeficiente de variación genética aditiva del diámetro es 12,42%. Coeficientes de variación genética próximos al 10% indican ganancias sustanciales por selección (López, 2004).

Las estimaciones de ganancias logradas promediando directamente las estimaciones BLUP de los Valores de Mejora (White and Hodge, 1989) y permiten escoger el esquema de selección más conveniente. La selección de las 20 familias que muestran los mejores desempeños en diámetro promueve ganancias del 5,48%.

Modelos de estimación

En la Tabla 4 se muestran los ajustes de los modelos de estimación BCA y espaciales probados. El cambio del logaritmo de la verosimilitud entre el modelo BCA y

espacial es significativo cuando se incluye a las columnas como efecto aleatorio (LRT=7,4). El valor del error estándar de la diferencia también señala este modelo como el más adecuado.

El desempeño de las 20 mejores familias del ensayo se muestra en la Figura 3. Estas estimaciones logradas por BLUE utilizando a las familias como efectos fijos del modelo son claramente diferentes de los promedios predichos con BLUP. Las estimaciones se refieren al desempeño de las familias en el ensayo y las predicciones a su desempeño futuro.

Estas estimaciones fueron usadas tradicionalmente para la selección de variedades en los experimentos de mejoramiento genético. Sin embargo, los

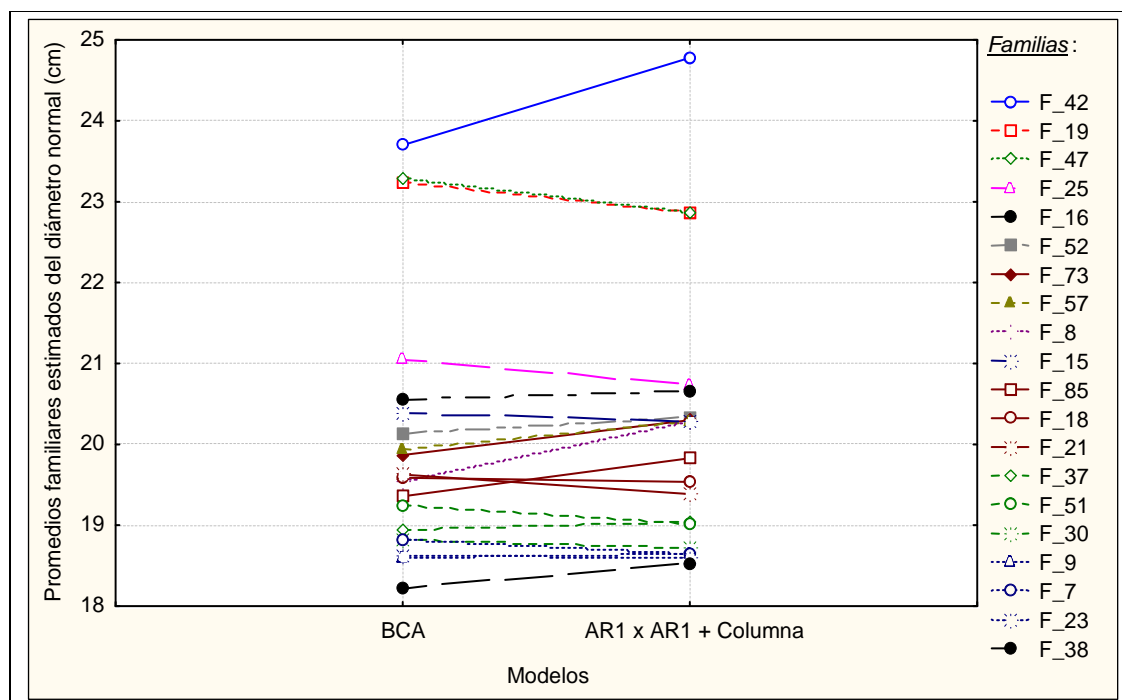


Figura 3. Comparación de promedios familiares estimados por los modelos

parámetros genéticos se consideran variables aleatorias y exigen el tratamiento correspondiente. El advenimiento de los modelos mixtos permitió su tratamiento adecuado.

IV. Conclusiones

El análisis espacial contribuye a aumentar la precisión en la predicción y estimación de parámetros en experimentos genéticos.

La modelación de los residuos espacialmente correlacionados permitió controlar la variación ambiental.

El modelo espacial que incluye las columnas como efecto aleatorio es más eficiente que el modelo BCA para la selección familiar.

Recomendaciones

El abordaje espacial en el análisis de datos no obvia la necesidad de establecer los ensayos con diseños experimentales adecuados que evitan que el efecto de los tratamientos se confunda con el efecto ambiental.

Referencias Bibliográficas

- Apiolasa, L. (2006). ASReml cookbook, Retrieved 18, February 2006. From <http://uncronopio.org/ASReml/>.
- Borrvalho, N. (1998). Genetics parameters estimation, En: Curso Mejora Genética Forestal Operativa. Editores: Ipinza R.; Gutierrez B.; Emhart V. Valdivia, Chile. 422 pp.
- Costa e Silva, J. Dutkowski, G. and Gilmour A. (2001), En: Analysis of early treeheight in forest genetic trials is enhanced by including a spatially correlated residual Can. J. For. Res. 31: 1887–1893.
- Dutkowski, G.; Costa e Silva, J; Gilmour, A. (2002). Spatial analysis methods for forest genetic trials. Can. J. For. Res. 32: 2201 – 2214.
- Falconer, D. (1981). Introduction to Quantitative Genetics. Longman & Co., New York, NY. 340p.
- Gezán Pacheco, S.; Torres Cuadros, J. (1998). Metodologías para la Determinación del Valor de Mejora. En: Mejora Genética Forestal Operativa. Editores: Ipinza, Gutiérrez, Emhart. Valdivia, Chile. Pp. 311-337.
- Gilmour, A.R., Cullis, B.R., and Verbyla, A.P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. J. Agric. Biol. Environ. Stat. 2: 269–293.
- Henderson, C.R. (1984). Applications of Linear Models in Animal Breeding. University of Guelph. Ontario, U.S.A. 462 pp.
- Ipinza Carmona, R. (1998). Diseño de Ensayos Genéticos. En: Mejora Genética Forestal Operativa. Editores: Ipinza, Gutiérrez, Emhart, Valdivia. Chile. Pp. 249-280.
- López, C. (2004). Variación genética en procedencias y progenies de *Eucalyptus camaldulensis* introducidas en el Noroeste Argentino. Tesis Doctoral. 89 pp.
- Mora Garcés, A. y Ramírez, G. (2000). Consideraciones prácticas sobre el análisis estadístico espacial de experimentos forestales de campos. Revista Forestal Venezolana, 44(2) 21-28 p
- Rocha, M.G.B.; Pires, I.E.; Xavier, A.; Cruz, C.D.; Rocha, R.B. (2006). Avaliação Genética de Progenies de Meio-Irmãos de *Eucalyptus urophylla* Utilizando os Procedimentos REML/BLUP e E(QM). Ciência Florestal, Santa Maria, vol. 16, Nº 4, Pp. 369-379.
- Searle, S. R. (1971). Linear models. John Wiley & Son, New York. 532 pp.
- Valera, F.P.; Kageyama, P.Y. (1995). Parámetros Genéticos y Espaciamento en Progenies de *Eucalyptus saligna* SMITH. Revista IPEF Nº 48/49. Brasil. Pp. 61-76.
- White, T. L. and Hodge, G.R. (1989). Prediction Breeding Values with Applications in Forest Tree Improvement. Kluwer Academic Publishers. Dordrech, Netherlands. 367pp.