

AUTOCORRELACIÓN ESPACIAL: UN PROBLEMA COMÚN...MENTE OLVIDADO

Rafael Zas Arregui, Patricia Martínez Garbín y Raúl de la Mata Pombo

Centro de Investigaciones Ambientales CINAM-Lourizán. Apdo 127. 36080-PONTEVEDRA (España).
Correo electrónico: rzas.cifal@siam-cma.org

Resumen

Se presenta una revisión de la estructura espacial de distintas variables evaluadas en ensayos genéticos de *Pinus pinaster* en Galicia. La estructura espacial se analiza a partir de los semivariogramas residuales, gráficos que sintetizan la dependencia espacial de los residuos del modelo estadístico mediante la representación de la variación de la similitud entre árboles en función de la distancia que los separa. Se presentan los resultados para variables de crecimiento, forma, ramosidad y resistencia a enfermedades, analizadas a distintas edades en ensayos de progenie instalados bajo distintos diseños experimentales. En la mayoría de las situaciones los residuos no son espacialmente independientes. La intensidad del patrón espacial llega a intensidades sorprendentemente altas, especialmente para variables de crecimiento a edades avanzadas, que pueden disminuir la eficiencia estadística de los diseños experimentales. Se concluye que la autocorrelación espacial es la pauta general en los ensayos genéticos forestales de *P. pinaster*. El diagnóstico del requisito de independencia mediante la exploración de la estructura espacial de los residuos utilizando los semivariogramas residuales debería ser una práctica rutinaria a añadir a los diagnósticos comunes de los otros requisitos del análisis. En caso de detectarse un patrón espacial no aleatorio, el uso de técnicas de análisis espacial puede volverse imprescindible.

Palabras clave: *Diseño experimental, ANOVA, Requisito de independencia, Análisis espacial, Heterogeneidad ambiental, Pinus pinaster*

INTRODUCCION

El tamaño de los ensayos genéticos forestales es especialmente elevado debido al gran número de unidades genéticas evaluadas y al tamaño de las unidades experimentales (u.e.) necesario para que los árboles crezcan libres de competencia. Además, estos ensayos se suelen establecer en terrenos típicamente forestales, habitualmente irregulares y fuertemente heterogéneos. En estas condiciones, los diseños experimentales convencionales, como el diseño de bloques completos al azar (BCA) o incluso los

diseños de bloques incompletos (BI), pueden resultar insuficientes para absorber la variación ambiental. Los bloques del diseño resultan, con frecuencia, demasiado grandes e internamente heterogéneos. Cuando esta heterogeneidad se refleja en la variable dependiente, los datos presentan una autocorrelación espacial que supone la violación del requisito de independencia de los análisis estadísticos convencionales. Aunque este requisito raramente se comprueba, su incumplimiento conlleva importantes repercusiones que pueden alterar drásticamente las conclusiones derivadas de los análisis estadísticos

(ver ZAS, 2007 y referencias citadas allí). Los efectos de la autocorrelación espacial afectan tanto a la estimación de los componentes de la varianza y parámetros genéticos derivados, como a la predicción de los valores de mejora y a la precisión de éstas predicciones (ZAS, 2006a,b). La gravedad de estos efectos varía en función de la intensidad y escala del patrón espacial en relación al tamaño de los bloques del diseño experimental (DE LA MATA et al., 2008).

Muchos de los programas de estadística modernos incluyen nuevos métodos que permiten bien incorporar la estructura espacial de los datos en los propios modelos (e.g. LITTELL et al., 1996), o bien ajustar o corregir la variable dependiente eliminando la autocorrelación espacial de ésta (e.g. ZAS, 2006b). Son muchos los trabajos que en el campo de la genética forestal, han demostrado en los últimos años las ventajas de estos métodos sobre los métodos convencionales (e.g. DUTKOWSKI et al., 2006; ZAS, 2006b).

Aunque la autocorrelación espacial es un fenómeno común en los ensayos genéticos forestales, no todas las variables ni todos los sitios ensayos padecen este problema. En ocasiones, por ejemplo, el establecimiento de diseños experimentales latinizados o con bloques de pequeño tamaño puede resultar suficiente para absorber la mayoría de la variación espacial (DE LA MATA et

al., 2008). En cualquier caso, el conocimiento de la estructura del patrón espacial en una colección grande de ensayos puede ser de gran utilidad para la optimización de los diseños experimentales de nuevos ensayos y/o para la formulación de los modelos de análisis espacial más apropiados (FU et al., 1999). Con el fin de tener una panorámica general de la incidencia de la autocorrelación espacial, en el presente trabajo se presenta una revisión de la estructura espacial de distintas variables evaluadas a varias edades en distintos ensayos de progenie de *Pinus pinaster* establecidos dentro del programa de mejora genética de la especie para Galicia.

MATERIAL Y METODOS

Los datos utilizados en el presente trabajo provienen de evaluaciones realizadas a edades de entre 1 y 8 años en 15 ensayos genéticos de *P. pinaster*. El material de todos los ensayos se trata de familias de polinización abierta procedentes de árboles plus seleccionados en la procedencia 1a-Noroeste Litoral y presentes en los huertos semilleros clonales de primera generación existentes en Galicia. Los ensayos provienen de 4 series diferentes. La primera (ASN, CRT, BAM y LAL), establecida en 1994-95, consta de 4

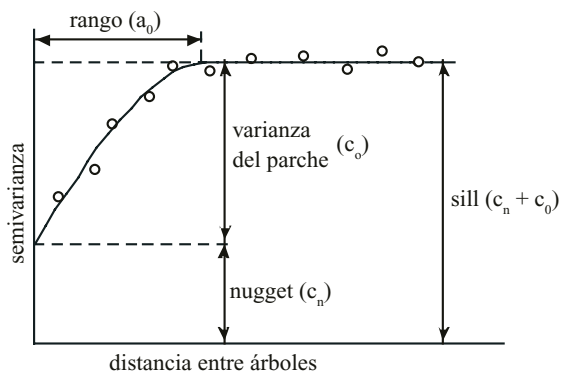


Figura 1. Esquema de un semivariograma hipotético mostrando el semivariograma empírico (puntos) y el modelo teórico ajustado (línea). Se indican los tres parámetros que caracterizan la estructura espacial: el rango (a_0) que es la distancia a la que se pierde la dependencia espacial e indica la escala del patrón, el nugget (c_n) o semivarianza a distancia cero que representa la varianza debida a errores de muestreo y/o ocasionada por la dependencia espacial a escalas inferiores a las muestreadas, y la varianza del parche (c_o) que refleja la intensidad de la dependencia espacial. La asintota o sill ($c_n + c_o$) es un estimador de la varianza total de la variable. La intensidad del patrón espacial viene dada por el cociente entre la varianza del parche y esta asintota ($c_o / (c_o + c_n)$)

ensayos de progenie con 78-111 familias, establecidos bajo un diseño experimental de BCA con 10 bloques y u.e. de 5 plantas en línea. En la segunda serie, establecida en el año 2000, sólo hay información de un único ensayo de progenie (LAR) con 116 familias y un diseño BCA con 25 bloques y u.e. monoárbol. La tercera serie se trata de tres ensayos factoriales familia x fertilización (FxF) (CEA, REB2 y RIA) establecidos en 2003 bajo un diseño split-plot con 10 bloques, 9 tratamientos de fertilización (factor principal), y 31 familias (factor secundario) en u.e. monoárbol. Finalmente, la última serie consiste en 7 ensayos de progenie (COR, CHA, FOR, FRI, MAN, NOC y REB) con 116 familias y un diseño de BI resolubles con 96 bloques de 10 familias cada uno agrupados en 8 réplicas o bloques completos, y u.e. de 3 plantas contiguas. El tamaño de los bloques en las cuatro series es de aproximadamente 4500, 1100, 1800 y 180 m², respectivamente. Detalles de estos ensayos pueden consultarse en ZAS (2006b), ZAS et al. (2007), MARTINS et al. (2008) y DE LA MATA et al. (2008), respectivamente.

Las variables consideradas fueron el crecimiento en altura y diámetro a distintas edades, variables de forma y ramosidad (rectitud del fuste en escala 1-6, número de verticilos y ángulo de inserción de ramas en escala 1-4), variables relacionadas con la resistencia a distintas patologías (sintomatología de la infección causada por el hongo *Fusarium oxysporum* en escala 0-5, daños provocados por el curculiónido descortezador *Hylobius abietis* en escala 0-30, y mortalidad generada por el patógeno del suelo *Armillaria ostoyae*) y variables relacionadas con la eficiencia del uso de nutrientes (concentración foliar de nitrógeno y fósforo).

La posición relativa de cada planta se determinó mediante una estación total o utilizando fotografías aéreas ortográficas a escala ~1:1000. La estructura espacial de cada una de estas variables se exploró mediante la construcción de los semivariogramas de los residuos de éstas tras eliminar el efecto familiar, y el correspondiente a la fertilización y fertilizaciónxfamilia en el caso de los tres ensayos factoriales FxF. Estos residuos se obtuvieron aplicando un ANOVA de una vía (o ANOVA factorial en el caso de los ensayos FxF) utilizando el procedimiento MIXED de

SAS y considerando todos los factores fijos. Con el fin de evitar discontinuidades en el patrón espacial alrededor de las fronteras de los bloques, el efecto bloque no se tuvo en cuenta en estos análisis. De esta manera, el patrón espacial de estos residuos representa el patrón espacial de la variable en sentido amplio. Parte de esta variación espacial podrá ser absorbida por los respectivos diseños experimentales y por tanto, el hecho de que los residuos presenten una estructura espacial no aleatoria no implica directamente la violación del requisito de independencia en un análisis que respete el diseño original.

La estructura espacial de los residuos se exploró mediante la construcción de un semivariograma (Figura 1). Los semivariogramas son figuras muy simples que sintetizan la estructura espacial mediante la representación de la variación de la similitud entre árboles en función de la distancia que los separa (GALLARDO, 2006). La función representada en un semivariograma es la semivarianza, que es una medida de dispersión entre los valores observados en dos puntos. Esta función se calcula para todos los posibles pares de árboles en cada ensayo, representando los valores promedios agrupando los pares según la distancia que separa los puntos del par. Para una determinada clase de distancia, la semivarianza $\gamma(h)$ se calcula como:

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n [z(s_i) - z(s_{i+h})]^2$$

siendo n el número de pares separados por la distancia h , $z(s_i)$ el valor residual en un árbol situado en la posición s_i y $z(s_{i+h})$ el valor de un árbol situado a una distancia h de s_i . Si la distribución de los datos es espacialmente independiente, el semivariograma resultará esencialmente plano. Por el contrario, si hay dependencia espacial, la semivarianza es menor a distancias cortas, aumenta a distancias intermedias y alcanza típicamente una asíntota (*sill*) a distancias mayores (Figura 1). La forma del semivariograma ofrece, a su vez, información sobre la estructura espacial de los datos. Si la semivarianza aumenta de forma continua con la distancia y no se alcanza nunca una asíntota, la estructura espacial corresponde a un gradiente (modelo lineal). Por el contrario, cuando se alcanza la asíntota (modelo esférico) la estructura espacial corresponde a un

parcheado, cuya escala o tamaño de parche (rango) viene definida por la distancia a la cual se alcanza la asíntota (Figura 1). Para cuantificar el grado y escala del patrón espacial es necesario ajustar una función matemática que permita caracterizar en unos pocos parámetros la estructura espacial de los datos. Los modelos más utilizados son los ya comentados lineal y esférico, y el exponencial en el que la asíntota se alcanza de forma exponencial (GALLARDO, 2006). Estas funciones se pueden ajustar mediante procedimientos de regresión no lineal como el procedimiento NLIN de SAS. El coeficiente de determinación (r^2) de este ajuste es un indicativo del grado de definición de la estructura espacial. A su vez, para estructuras en parche, el cociente entre la varianza del parche y la varianza *sill* es un indicativo de la intensidad del patrón espacial, reflejando qué porcentaje de la variación se debe a la estructura espacial (ver Figura 1).

RESULTADOS Y DISCUSION

La estructura espacial para variables de crecimiento se resume en la tabla 1. En todos los sitios de ensayo salvo en uno y para todas las edades evaluadas, los elevados coeficientes de determinación ($r^2 > 0.90$) indican una estructura espacial no aleatoria fácilmente modelizable mediante alguno de los modelos típicos de geoestadística. La estructura espacial varía desde patrones en gradiente hasta patrones parcheados, siendo los primeros más frecuentes a edades tempranas. En los casos de patrones en manchas, el tamaño de éstas (rango efectivo) es ampliamente variable, desde escasos 20 m hasta más de 300 m, con un promedio de aproximadamente 65 m. Este último valor es claramente mayor que el presentado por FU et al. (1999) para una colección de ensayos de *Pseudotsuga menziesii* (rango promedio = 18 m), si bien estos

Ensayo	Altura 1 año				Altura 2 años				Diámetro 2 años			
	modelo	a_0^1 (m)	I(%)	r^2	modelo	a_0^1 (m)	I(%)	r^2	modelo	a_0^1 (m)	I(%)	r^2
CHA	lin			0,96	exp	35	32	0,96				
COR	sph	35	25	0,94	sph	35	24	0,93				
FOR	lin			0,97	lin			0,98				
FRI	exp	21	26	0,92	exp	30	32	0,99				
MAN	exp	43	28	0,97	exp	28	46	0,99				
NOC	sph	23	32	0,97	sph	32	43	0,98				
REB	lin			0,98	lin			0,92				
RIA	sph	17	11	0,95	sph	23	17	0,90	sph	22	21	0,94
CEA	sph	28	42	0,93	sph	31	55	0,95				
REB2	sph	22	16	0,79	sph	24	23	0,94	sph	57	28	0,96
	Altura 3 años				Altura 4 años				Diámetro 3 años			
RIA	sph	34	22	0,95	sph	34	29	0,96	sph	96	24	0,94
CEA	sph	31	53	0,97								
REB2	sph	27	26	0,79	sph	35	29	0,73	sph	69	25	0,90
CRT	sph	117	34	0,99								
	Altura 5 años				Altura 8 años				Diámetro 8 años			
ASN					exp	20	34	0,99	exp	21	25	0,99
BAM					exp	24	54	0,99	exp	13	46	0,99
CRT	sph	110	35	0,98	exp	81	50	0,99	exp	101	42	0,99
LAL	exp	26	36	0,99	exp	30	49	0,99	exp	28	38	0,98

Tabla 1. Estructura espacial del crecimiento en altura y en diámetro evaluado a distintas edades en distintos ensayos de progenie del programa de mejora de *P. pinaster* en Galicia. Se presenta el mejor modelo de ajuste (lin = lineal, exp = exponencial, sph = esférico), el rango (a_0), la intensidad del patrón espacial (I%) y el coeficiente de determinación del ajuste del modelo teórico. ¹El rango efectivo del modelo exponencial (distancia a la que se alcanza el 95% del sill) es $3 \times a_0$.

autores modelizaron la estructura espacial una vez eliminadas las tendencias o gradientes (*detrended residuals*) y por tanto su modelización se centra en escalas menores. En nuestro caso, la estructura espacial a pequeña y gran escala se modeliza simultáneamente, resultando en rangos mayores. Es por ello también, que en muchas situaciones, los modelos que mejor se ajustan son los modelos lineales o exponenciales, mientras modelizando residuos sin tendencias, el modelo que mejor se ajusta suele ser el esférico. En cualquier caso, el tamaño del parche es en muchas situaciones menor que el tamaño de los bloques del diseño experimental, lo que indica una forzosa heterogeneidad intra-bloque y, por tanto, una escasa eficiencia de los diseños. Incluso en los diseños de BI, el menor tamaño de los bloques es, en ocasiones, incapaz de absorber la dependencia espacial a menor escala (DE LA MATA *et al.*, 2008). La falta de eficiencia de los diseños se traduce en estimaciones erróneas o sesgadas de varianzas, parámetros genéticos y valores de mejora, siendo fundamental en estos casos recurrir a técnicas de análisis espacial que incorporen la autocorrelación de los datos en el análisis (ver ZAS, 2006b, 2007; ZAS *et al.*, 2007).

La intensidad del patrón varía entre a penas un 10% y más del 50%, siendo patente una tendencia a intensidades más elevadas con la edad. Esto último puede verse fácilmente al comparar la estructura espacial en una de las parcelas a distintas edades (Figura 2). A medida que la masa crece, tanto el rango como la intensidad del

patrón espacial va aumentando. Esto es típico que ocurra, y en general el reflejo de la heterogeneidad ambiental en el crecimiento tiende a aumentar con la edad hasta que la competencia entre árboles genera una autocorrelación negativa que contrarresta la anterior positiva (MAGNUSSEN, 1994).

La estructura espacial para el resto de variables analizadas se presenta en la tabla 2. Aunque la presencia de estructuras no aleatorias vuelve a ser la tónica general se observan coeficientes de determinación de los modelos teóricos más bajos e intensidades del patrón claramente menores. Es destacable que ciertos caracteres evaluados de forma subjetiva, como la rectitud del tronco, pueden mostrar autocorrelación espacial positiva como consecuencia de la inevitable variación del criterio subjetivo del evaluador. Así, no es infrecuente que en determinados momentos (y por tanto en determinadas posiciones) el observador tienda a valorar más alto o más bajo la variable en cuestión.

La incidencia de enfermedades o plagas es otro de los típicos casos donde la irregularidad espacial de la infección natural en un ensayo de campo puede comprometer el análisis de la variación genética en resistencia. Así, las enfermedades fúngicas tienden a expandirse desde los focos de infección generando las correspondientes estructuras de parches. En el caso de *F. oxysporum* y *A. ostoyae*, esta dependencia espacial no es absorbida por los bloques del diseño experimental, siendo necesario recurrir a métodos de análisis espacial para un correcto análisis (MARTÍNS *et*

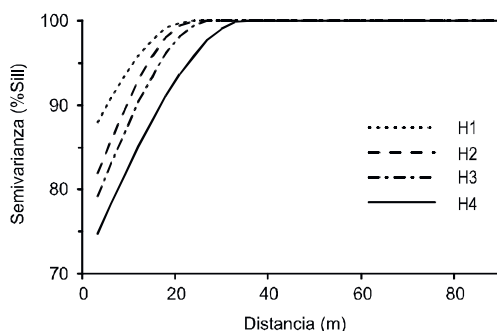


Figura 2. Evolución del patrón espacial para el crecimiento en altura en el ensayo familia x fertilización de Rebordelo durante los primeros 4 años de edad. Los respectivos semivariogramas teóricos (modelo esférico) se han normalizado representado la semivarianza en función del Sill, en porcentaje

		Edad	modelo	a_0^1 (m)	I (%)	R ²
<i>Variables de forma y ramosidad</i>						
Rectitud del fuste evaluada en escala de 6 niveles	ASN	8	exp	46	12	0,78
	BAM	8	exp	21	17	0,88
	CRT	8	exp	31	18	0,97
	LAL	8	exp	7	24	0,83
Número de verticilos	ASN	8	exp	21	10	0,93
	BAM	8	exp	14	14	0,90
	CRT	8	exp	47	30	0,97
	LAL	8	exp	20	8	0,97
Ángulo de inserción de ramas (escala de 4 niveles)	ASN	8	semivariograma plano			
	BAM	8	exp	9	19	0,93
	CRT	8	exp	7	9	0,94
	LAL	8	semivariograma plano			
<i>Resistencia a enfermedades</i>						
Síntomatología por infección de <i>F. oxysporum</i>	CEA	2	sph	29	35	0,96
	CEA	3	sph	106	42	0,98
Supervivencia por ataque de <i>A. ostoyae</i>	LAR	3	exp	21	12	0,94
Mordeduras por ataque de <i>H. abietis</i>	RIA	1	semivariograma plano			
	RIA	2	semivariograma plano			
<i>Concentración de nutrientes</i>						
Concentración foliar de nitrógeno	REB2	2	semivariograma plano			
Concentración foliar de fósforo	REB2	2	sph	48	32	0,68

Tabla 2. Estructura espacial de variables de forma y ramosidad, variables relacionadas con la resistencia a enfermedades, y variables relacionadas con la eficiencia nutricional evaluadas a distintas edades en ensayos de progenie de *P. pinaster* en Galicia. Se presenta el mejor modelo de ajuste (lineal, exponencial o esférico), el rango (a_0), la intensidad del patrón espacial (I%) y el coeficiente de determinación del ajuste del modelo teórico. ¹ El rango efectivo del modelo exponencial (distancia a la que se alcanza el 95% del sill) es $3 \times a_0$.

al., 2008; ZAS et al., 2007). Por el contrario, el tamaño del ensayo es suficientemente pequeño para que el ataque de *H. abietis*, que se aproxima a la zona volando desde largas distancias, se muestre de forma uniforme (ZAS et al., 2006).

Los resultados obtenidos para las concentraciones de nutrientes coinciden con lo esperado y reflejan las ventajas de los sitios de buena calidad de estación para el establecimiento de ensayos genéticos. Así, tal como observaron GALLARDO Y COVELO (2005) en *Quercus robur*, el patrón espacial es evidente para el nutriente limitante (fósforo) pero no para el excedentario (nitrógeno). La irregularidad en la disponibilidad de fósforo se plasma con facilidad en el fósforo captado, reflejándose esta variación en el crecimiento de las plantas, claramente limitado por la escasez de este nutriente.

CONCLUSIONES

Se concluye que la autocorrelación espacial es la pauta general en los ensayos genéticos forestales de *P. pinaster*. Pudiendo ésta implicar la violación del requisito de independencia de los análisis estadísticos, es fundamental explorar siempre la estructura espacial de los datos y en caso de detectarse un patrón espacial no aleatorio, recurrir al uso de técnicas de análisis espacial.

BIBLIOGRAFÍA

DE LA MATA, R.; ZAS, R. Y MERLO, E.; 2008. Control de la autocorrelación espacial mediante diseños experimentales y métodos de análisis espacial en ensayos de progenie

- de *Pinus pinaster*. *Cuad. Soc. Esp. Cienc. For.* 24: 33-38.
- DUTKOWSKI, G.W.; COSTA-E-SILVA, J.; GILMOUR, A.R.; WALLENDORF, H. & AGUIAR, A.; 2006. Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. *Can. J. For. Res.* 36(11): 1851-1870.
- FU, Y.B.; YANCHUK, A.D. & NAMKOONG, G.; 1999. Spatial patterns of tree height variations in a series of Douglas-fir progeny trials: implications for genetic testing. *Can. J. For. Res.* 29(6): 714-723.
- GALLARDO, A.; 2006. Geoestadística. *Ecosistemas* 3: <http://www.revistaecosistemas.net>.
- GALLARDO, A. & COVELO, F.; 2005. Spatial pattern and scale of leaf N and P concentration in a *Quercus robur* population. *Plant Soil* 273(1-2): 269-277.
- LITTELL, R.C.; MILLIKEN, G.A.; STROUP, W.W. & WOLFINGER, R.D.; 1996. *SAS System for mixed models*. SAS Institute. Cary, NC.
- MAGNUSSEN, S.; 1994. A method to adjust simultaneously for spatial microsite and competition effects. *Can. J. For. Res.* 24(5): 985-995.
- MARTÍNS, P.; MOREIRA, X.; ZAS, R.; SAMPEDRO, L. Y SOLLA, A.; 2008. Variación genética y efecto de la fertilización en la susceptibilidad de *Pinus pinaster* a *Fusarium oxysporum*. *Cuad. Soc. Esp. Cienc. For.* 25: en prensa.
- ZAS, R.; 2006a. Consecuencias de la estructura espacial de los datos en el diseño y análisis de experimentos en campo. *Ecosistemas* 3: <http://www.revistaecosistemas.net>.
- ZAS, R.; 2006b. Iterative kriging for removing spatial autocorrelation in analysis of forest genetic trials. *Tree Genet. & Genomics* 2(4): 177-186.
- ZAS, R. 2007. Autocorrelación espacial y el diseño y análisis de experimentos. En: F. Maestre, A. Escudero y A. Bonet (eds.), *Introducción al análisis espacial de datos en ecología y ciencias ambientales: métodos y aplicaciones*. U. Rey Juan Carlos, AEET y CAM. Madrid (en prensa).
- ZAS, R.; SAMPEDRO, L.; PRADA, E.; LOMBARDEO, M.J. & FERNÁNDEZ-LÓPEZ, J.; 2006. Fertilization increases *Hylobius abietis* L. damage in *Pinus pinaster* Ait. seedlings. *Forest Ecol. Manage.* 222(1-3): 137-144.
- ZAS, R.; SOLLA, A. & SAMPEDRO, L.; 2007. Variography and kriging allow screening *Pinus pinaster* resistant to *Armillaria ostoyae* in field conditions. *Forestry* 80(2): 201-209.