

# MODELO INDIVIDUAL DE PRODUCCIÓN DE PIÑÓN DE *PINUS PINEA* L. COMO CRITERIO DE SELECCIÓN FENOTÍPICA

J. GORDO ALONSO<sup>(1)</sup>, S. MUTKE REGNERI<sup>(2)</sup>, L. GIL SANCHEZ<sup>(2)</sup>

<sup>(1)</sup> Junta de Castilla y León, Servicio Territorial de Medio Ambiente  
C/Duque de la Victoria 5, 47001 Valladolid

<sup>(2)</sup> ETSIM Madrid, Unidad de Anatomía, Fisiología y Genética Forestal  
Ciudad Universitaria s/n, 28040 Madrid

S.MUTKE@forestales.org

## RESUMEN

Se discute la metodología de la selección masal de árboles sobresalientes en producción de piña en la región de procedencia Meseta Norte del pino piñonero (*Pinus pinea*) para constituir la primera generación de la población de mejora. Para su selección, se ha elaborado un modelo de producción individual de piñón, partiendo de los datos de cosecha de una colección de los árboles candidatos, aforada durante ocho años para determinar en cada rodal de muestreo los mejores fenotipos. Como variables de ponderación de la productividad se han incluido una constante del rodal, el tamaño del árbol, estimado por su área basimétrica, y su espacio vital disponible, estimado por distancias árbol-árbol. El modelo logarítmico lineal obtenido engloba la mayor parte de la variación observada entre árboles, mientras los valores residuales respecto la superficie de respuesta del modelo se interpretan como estimadores de la calidad productiva del candidato independiente de las variables incluidas en el modelo.

**P.C.:** pino piñonero, selección de árboles plus, modelo de producción de piñón

## SUMMARY

Methodology of cone-bearing plus tree selection in stone pine (*Pinus pinea*) forests in Central Spain is discussed, based on eight kernel yields. An individual kernel yield model was established in order to weigh yield-influencing parameters of each tree. The resulting log-linear model, which includes a sample-plot constant, the DBH and the growing-area of the tree, achieves a coefficient of determination about 0,8. Residual kernel-yield proportion not explained by the model's response surface is used as selection criterion.

**K.W.:** *Pinus pinea*, plus tree selection, individual kernel-yield model

## INTRODUCCION

La región de procedencia Meseta Norte del pino piñonero, *Pinus pinea* L., comprende las masas de la especie situadas en la Tierra de Pinares vallisoletana, segoviana y abulense que destacan por la importancia del piñón en la economía y cultura rural de esta comarca. Tradicionalmente, este fruto seco se ha obtenido como recurso forestal de las masas naturales de la especie. Actualmente, el fin previsible a medio plazo de la agricultura de los cultivos tradicionales en secano, mantenida sólo por las ayudas públicas, dejará ociosas en la Meseta Castellana amplias superficies de tierra agrícola, para las que será interesante encontrar usos alternativos. Como un cultivo con futuro se presentan las plantaciones de pino piñonero injertado destinadas a la producción de piñón, estando actualmente en fase de investigación y desarrollo las mejoras vegetales y tecnológicas necesarias que permiten al propietario privado optimizar su rentabilidad.

Concretamente, dentro del programa de mejora genética del pino piñonero en Castilla y León, una de las líneas de trabajo desarrolladas se dirige hacia la obtención de genotipos grandes productores de piña, para su propagación en estas futuras plantaciones injertadas. En la presente comunicación se discute el método de evaluación de los candidatos en monte, cuya cosecha de piña y piñón se ha medido durante 8 años al objeto de determinar los fenotipos realmente sobresalientes durante más de un ciclo completo de vecería. Se ha de tener en cuenta que la producción de fruto no permanece en el árbol, por lo que es necesario volver cada año para su aforo, para obtener la misma información de producción plurianual que en el caso de la madera, que sí se acumula año tras año en el tronco y se mide en pocos minutos mediante una simple muestra de barrena de Pressler.

Como antecedente de la evaluación de producción individual de piña, cabe citar el trabajo de Cappelli (1958), que encontró para el piñonero dentro de rodales coetáneos unas regresiones lineales de la producción de piña con el tamaño del árbol, estimado por su diámetro normal (DAP) o el diámetro de copa. Basado en la relación encontrada, Cappelli y algunos autores posteriores utilizan el valor de la cosecha referida al diámetro o la superficie de copa (SC) como criterio de selección

fenotípica. Se trata de una selección según el método de línea base, al elegir los candidatos con el valor de producción por encima de la media correspondiente al valor de la variable de ponderación.

Otro parámetro con comprobada influencia sobre la abundancia individual de la fructificación es la densidad de la masa, ya que un mayor espacio vital y acceso a los recursos (agua, luz) permiten al árbol una mayor floración y fructificación. El criterio de selección podría ser en este caso por analogía la producción por metro cuadrado de espacio vital (EV).

No obstante, la comparación de este método con otro basado en la regresión matemática de la producción contra la variable independiente (DAP, SC o EV) muestra que el primero está sesgado en contra del tamaño del árbol, ya que se basa en el ángulo polar del punto en el sistema de coordenadas (variable independiente; producción) en comparación con el ángulo medio. Existe una pérdida de información al ignorar la otra componente de la coordenada polar del punto, la distancia al origen, lo que favorece a los árboles menores al exigirles menor diferencia sobre la línea base para ser elegidos. Además, la línea base es siempre recta y pasa obligatoriamente por el origen (0;0).

Se ofrece como alternativa evidente la regresión por mínimos cuadrados como método insesgado. Carece de las limitaciones del método “angular”, ya que la línea base no será obligatoriamente lineal (por ejemplo, puede depender de manera lineal del diámetro de la copa, o cuadrático (superficie), o cúbico (volumen), logarítmico, etc.), ni tiene porqué pasar por el origen (puede haber una parte de la copa no relacionada con la fructificación). Como criterio de evaluación fenotípica sirve la variación no explicada por las variables independientes del modelo, es decir, la producción por encima de la esperada del árbol según la ecuación de regresión. Para que la selección basada en el modelo sea válida, será necesario que se cumpla la homocedasticidad o independencia de la varianza frente a las variables.

Por último, la regresión permite incluir más de una variable, estimando la contribución de cada una de ellas en el valor de la producción. Un modelo de regresión múltiple con  $n$  fuentes de variación cuantitativas, que análogo a una línea base en dos dimensiones representa un hiperplano de  $n$  dimensiones en el espacio  $R^{n+1}$ , es el caso más simple de una superficie respuesta (*response surface*, Box & Wilson, 1951, cit. en Dean & Voss, 1999), que descompone el valor de la variable  $Y$  en la predicción del modelo  $\eta_i = E[Y_i]$  y un error o residuo  $\varepsilon_i$  de distribución normal y media cero:

$$Y_i = \eta_i + \varepsilon_i$$

Si la regresión múltiple es lineal para todas las variables y sin interacciones significativas, la superficie respuesta  $\eta$  es plana, mientras la inclusión de términos de mayor orden o de interacciones la curvará. Por otra parte, la introducción de variables cualitativas en el modelo corresponde a la introducción de términos constantes, y significa gráficamente que se proponen varias superficies paralelas entre sí, análogo a varias líneas de regresión paralelas en un sistema de coordenadas de dos dimensiones.

El objetivo de este trabajo es establecer un modelo de regresión para elegir el mejor árbol, o a lo sumo los dos mejores de cada rodal, que cuenta inicialmente con cinco candidatos, y constituir una selección de árboles *plus* de amplia base genética y mínimo parentesco para las siguientes fases del programa de mejora.

## MATERIAL Y METODOS

El material estudiado procede de la selección masal de árboles *plus*, en la que se marcaron en veinte rodales de buena producción de piña los cinco árboles aparentemente mejores, cien árboles en total, para seguir durante ocho años su producción de fruto (Mutke *et al.*, 2000). A la vista de los ciclos largos (6 a 10 años) e irregulares de vecería de la especie, este lapso de tiempo ha permitido abordar un ciclo completo desde el primer año de cosecha máxima relativa, con varios años muy malos y la recuperación del ciclo hasta otros dos años de cosecha buena. La metodología de la selección de los 100 árboles candidatos, las variables de rodal y árbol medidas y los datos de cosecha medidos durante ocho años, así como las correlaciones de las variables de producción de piña y piñón se publicaron en trabajos anteriores (Gordo, 1998; Mutke *et al.*, 2000), por lo que la presente comunicación se centra en el análisis estadístico del valor final de producción, la cosecha media anual de piñón (con cáscara) de cada árbol candidato en ocho años,  $P$ . Como punto de partida, se dispone de datos completos de ocho cosechas de sólo 74 árboles, al haberse perdido datos de cosecha o extracción de piñón de algún año de los restantes candidatos. Para la evaluación de éstos se ha

repetido el mismo método expuesto en la presente para ocho años con el conjunto de datos excluyendo el año de cosecha incompleta (no incluido en la comunicación).

Con el fin de construir mediante el procedimiento matricial GLM un modelo de regresión lineal múltiple de la esta variable, se han tenido en cuenta las siguientes variables como posibles fuentes de variación de la cosecha:

Variable cualitativa

R (rodal o punto de muestreo con 5 candidatos en una superficie de unas 4 hectáreas. Esta variable engloba la calidad de estación)  $R \in \{1, 2, 3, \dots, 20\}$

Variables cuantitativas del árbol:

AB	área basimétrica	[m <sup>2</sup> ]
SC	superficie de proyección de copa	[m <sup>2</sup> ]
EV	espacio vital, determinado según distancia árbol-árbol	[m <sup>2</sup> ]
H	altura total	[m]
Edad		[años]

En la distribución de la variable *P* de los 74 árboles con datos completos, se observó una asimetría positiva fuerte, debida a la producción muy superior a la media de unos pocos árboles. Se relaciona con el mayor tamaño del árbol (en área basimétrica o superficie de copa) o con un mayor espacio vital disponible (estimado a partir de la distancia al quinto árbol más cercano, como la inversa de la densidad de la parcela circular imaginaria correspondiente).

Además, al analizar gráficamente la correlación de la producción de piñón *P* con cada una de las variables independientes, se apreció una falta de homogeneidad de la varianza, al abrirse la nube de puntos hacia la derecha, conforme aumenten los valores de las variables independientes; tampoco hubo homocedasticidad entre rodales. Para resolver tanto la falta de normalidad en la distribución de la variable respuesta como su heterocedasticidad, por constituir esta última una violación de los supuestos básicos para poder analizar la varianza de la variable (Dean & Voss, 1999), se utilizó la transformación por el logaritmo natural (Mead, 2000):

$$LNP = \ln(P+1)$$

Los gráficos respecto a esta variable transformada indicaron un comportamiento satisfactorio; la distribución de las observaciones de la variable LNP cumple los tests de normalidad. El modelo para el logaritmo de la producción media anual de un árbol *i* dentro del rodal *R* incluirá una constante *a*<sub>0</sub>, un efecto aditivo del rodal *λ*<sub>*R*</sub>, y los efectos lineales de *n* variables cuantitativas transformadas por su logaritmo, quedando sin explicar el valor residual *ε*<sub>*i*</sub>, según la ecuación:

$$LNP_{i(R)} = a_0 + \lambda_R + a_1 \ln V_1 + a_2 \ln V_2 + \dots + a_n \ln V_n + \epsilon_i$$

Este modelo log-lineal corresponde a un modelo linealizado que evalúa efectos aditivos a escala logarítmica, lo que equivale a efectos multiplicativos a escala natural (Mead, 2000), ya que al deshacer el logaritmo se transforma en:

$$P+1 = e^{a_0 + \lambda_R} V_1^{a_1} V_2^{a_2} \dots V_n^{a_n} e^{\epsilon}$$

**RESULTADOS Y DISCUSION**

En el análisis gráfico y por regresiones simples de la variable de producción LNP con las variables transformadas de cada árbol, se encontraron significativas la variable nominal de rodal, *R*, y las variables LNG=ln(AB), LNS=ln(SC) y LNE=ln(EV), mientras el resto de las variables no aporta variación significativa a la producción de piñón de la muestra de árboles candidatos. De los posibles modelos que combinan las cuatro variables significativas, se han excluido aquéllos en los que coinciden la superficie de copa SC y el área basimétrica AB, por presentar una fuerte colinealidad. Por otra parte, en la construcción del modelo se han excluido en principio tres árboles con valores residuales excesivos que tenían una fuerte influencia sobre la estimación de los parámetros de la ecuación.

Según se refleja en la primera fila de la Tabla 1, al analizar primero por separado las relaciones de LNP con cada una de las variables consideradas, mayor efecto tiene el factor R del rodal, que absorbe el 76% de la varianza, con un  $r^2$  ajustado de 0,69. Sin embargo, este valor se debe sobre todo al efecto de tres rodales buenos y tres malos, mientras para los demás rodales el factor  $\lambda$  no difiere de cero a un nivel de significación del 95%. La importancia del factor rodal, que está influenciado por la calidad productiva de la estación, indica la dificultad de elegir árboles plus entre varios montes sin ponderar el rodal de origen de cada uno. Por esta razón, se estableció, paralelamente y desde el inicio de la fase de selección, un banco clonal que recoge material vegetativo de los 100 candidatos iniciales, para avanzar simultáneamente en la caracterización fenológica y productiva de estos genotipos en parcelas injertadas y el estudio de su variabilidad genética y fenotípica.

La segunda fuente de variación más importante es el tamaño del árbol, evaluado por el logaritmo de su área basimétrica LNG o de su proyección de copa LNS, con un  $r^2$  ajustado de 45% y 40% respectivamente. El logaritmo del espacio vital (LNE) por sí solo explicaría solamente un 5% de la variación de la variable respuesta. Los modelos de la segunda fila de la Tabla 1, que incluyen una relación logarítmica lineal de la producción con una de las variables cuantitativas más el factor fijo del rodal, mejoran el modelo basado sólo en este último sólo en un 5%, debido a que las diferencias de tamaño de árbol son mayores entre rodales que dentro de ellos. La inclusión de la variable adicional de espaciamiento LNE en los modelos de la última fila aporta otro 2%.

La calidad de los dos modelos de la última fila de la tabla es parecida al basarse en los mismos hechos: rodal, espaciamiento y tamaño del pino. Las siguientes consideraciones nos han llevado a dar preferencia al primero, basado en el área basimétrica. La determinación de la superficie de copa mediante la toma de cuatro radios de su proyección fue más inexacta que la toma del área basimétrica calculada a partir de la circunferencia del tronco, porque el tronco suelo asemejarse más a una figura circular que la copa, especialmente en masas con tangencia de copas. El modelo basado en el área basimétrica tiene además la ventaja de su medición más fácil y rápida, de interés para simplificar futuros trabajos de prospección en monte.

Además, el modelo que se basa en un coeficiente del rodal y los logaritmos LNG y LNE, representado a la izquierda de la última fila de la Tabla 1, requiere la exclusión de sólo dos en vez de tres árboles por valores en exceso aberrantes e influyentes, con lo que se establece para 72 árboles. El coeficiente  $a_1$  a escala logarítmica, o exponente del área basimétrica a escala real, es 0,563, es decir, cercano a la raíz cuadrada (diámetro normal), el coeficiente (exponente a escala real) del espacio vital  $a_2$  es 0,287, mientras que el efecto aditivo a escala logarítmica de los rodales ( $a_0 + \lambda_R$ ) se sitúa entre -2,18 y -4,04. El análisis de la varianza para la variable de respuesta LNP según el modelo elegido y su coeficiente de determinación quedan reflejados en la Tabla 2. En la Figura 1 se señalan los valores residuales de los árboles plus que han servido como criterio de su selección.

Este modelo explicaría más de 82% de la variación de la variable transformada de la producción LNP, lo que muestra la importancia de ponderar el valor de producción de cada árbol para evitar una selección sesgada. En caso contrario se corre el riesgo de elegir árboles muy productores en cifras absolutas sólo por ser muy grandes, o de árboles de producción alta por superficie de copa sólo por tener un copa reducida. Por otra parte, la influencia determinante del tamaño del árbol y de la competencia de la masa concuerda con los resultados obtenidos por Cañadas (2000), que estudia la producción individual de piña en pino piñonero del Sistema Central.

Es evidente que la componente de nuestro modelo que estima la competencia del árbol por el espacio vital no es satisfactoria, ya que la experiencia selvícola le atribuye un efecto bastante mayor sobre la fructificación. Si en vez de determinar solamente la distancia al quinto árbol más próximo, se hubiera medido el dato adicional del diámetro normal de los cinco árboles, o incluso el rumbo y la distancia de cada uno respecto al árbol central, es más que probable que se habría mejorado el modelo obtenido a través de índices de competencia individual. Pero debido a que el objetivo del estudio no es la elaboración de un modelo predictivo para la especie, sino la evaluación fenotípica de candidatos a árboles plus (el factor Rodal es efecto fijo, no aleatorio), no se prevé seguir en esta línea, que no obstante parece muy interesante para la investigación selvícola sobre la producción de fruto.

Por último, conviene plantearse el significado biológico y las limitaciones del valor residual como estimador del genotipo. Igual que en el factor cualitativo *Rodal* influyen la calidad de estación,

la constitución genética de los árboles, la edad y la estructura de la masa - razón por la cual absorbe buena parte de la variabilidad relacionada con los otros factores del modelo -, también en el tamaño del árbol participa su genotipo, por lo que la eliminación de este factor en la evaluación del árbol podría ser equivocada. Sin embargo, como árboles plus respecto a la producción de fruto se buscan precisamente los árboles con mayor tendencia a la fructificación independientemente de los demás factores, por lo que consideramos válido el planteamiento.

## CONCLUSIONES

La presente comunicación discute la conveniencia de estudiar mediante procedimientos matemáticos el efecto de variables del árbol individual sobre la variable objeto de selección fenotípica, su producción de fruto. Para el pino piñonero se comprueba la importancia de ponderar la cosecha de cada árbol por su tamaño y su situación de competencia con el arbolado circundante, y sobre todo por la productividad media de su rodal.

## BIBLIOGRAFIA

- CAÑADAS, N. (2000). *Pinus pinea L. en el Sistema Central (Valles del Tiétar y del Alberche): desarrollo de un modelo de crecimiento y producción de piña*. Tesis Doctoral, ETS Ingenieros de Montes, UPM, Madrid (sin publicar). 356 p.
- CAPPELLI, M. (1958). Note preliminari sulla produzione individuale di strobili in *Pinus pinea L. L'Italia Forestale e Montana*, anno XIII, n° 5: 181-203.
- DEAN, A. & VOSS, D. (1999). *Design and Analysis of Experiments*. Springer-Verlag, New York, Berlin, Heidelberg.
- GORDO, J. (1998). Programa de mejora genética de *Pinus pinea L.* en Castilla y León. *Montes* n° 52 (2-98): 71-84.
- MEAD, A. (2000). *An Introduction to the Analysis of Qualitative Data*. in: *Enhancing Breeding Processes of Fruit and Forest Woody Species: Experimental Design and Statistical Analysis*. CIHEAM-IAMZ Zaragoza: 73 pp.
- MUTKE, S.; GORDO, F. & GIL, L. (2000). *Selección de individuos de Pinus pinea L. grandes productores de fruto en masas de la Meseta Norte*. Actas (II) del Primer Simposio del Pino Piñonero (*Pinus pinea L.*) febrero 2000, Valladolid: 85-94.

## TABLAS Y FIGURAS

**Tabla 1. Comparación de los modelos de la producción media de 71 árboles candidatos**

<b>Modelo</b>	$LNP_i = a_0 + \lambda_R + \varepsilon_i$	$LNP_i = a_0 + a_1 LNG_i + \varepsilon_i$	$LNP_i = a_0 + a_1 LNS_i + \varepsilon_i$	$LNP_i = a_0 + a_1 LNE_i + \varepsilon_i$
<b>MSE (g.l.)</b>	0,09766 (54)	0,1839 (70)	0,2005 (70)	0,3162 (70)
<b>r<sup>2</sup></b>	0,769	0,449	0,399	0,053
<b>Modelo</b>	$LNP_i = a_0 + \lambda_R + a LNG_i + \varepsilon_i$	$LNP_i = a_0 + \lambda_R + a LNS_i + \varepsilon_i$	$LNP_i = a_0 + \lambda_R + a LNE_i + \varepsilon_i$	
<b>MSE (53 g.l.)</b>	0,07997	0,07122	0,08822	
<b>r<sup>2</sup></b>	0,814	0,835	0,795	
<b>Modelo</b>	$LNP_i = a_0 + \lambda_R + a_1 LNG_i + a_2 LNE_i + \varepsilon_i$		$LNP_i = a_0 + \lambda_R + a_1 LNS_i + a_2 LNE_i + \varepsilon_i$	
<b>MSE (52 g.l.)</b>	0,07221		0,06542	
<b>r<sup>2</sup></b>	0,835		0,851	

$\lambda_R$  - coeficiente del rodal R;  $LNP_i = \ln(P)$  - logaritmo de la producción media anual P del árbol i;  $LNG_i = \ln(1000AB)$  - logaritmo de su área basimétrica AB;  $LNS_i = \ln(SC)$  - logaritmo de la proyección de su copa SC;  $LNE_i = \ln(EV)$  - logaritmo de su espacio vital EV; MSE - error medio cuadrático; r<sup>2</sup> - coeficiente de regresión; g.l. - grados de libertad

**Tabla 2. Análisis de varianza (Tipo III) de LNP (72 árboles candidatos)**

Fuente de variación	S.C.	g.l.	C.M.	F	P
RODAL	8,26847	16	0,51678	6,52	0,0000 ***
LNG	1,16478	1	1,16478	14,70	0,0003 ***
LNE	0,60403	1	0,604025	7,62	0,0079 **
Residual	4,20016	53	0,0792482		

Total (Corr.)

23,3646

71

$r^2 = 0,8202$

$r^2$  (ajustado para g.l.) = 0,7592

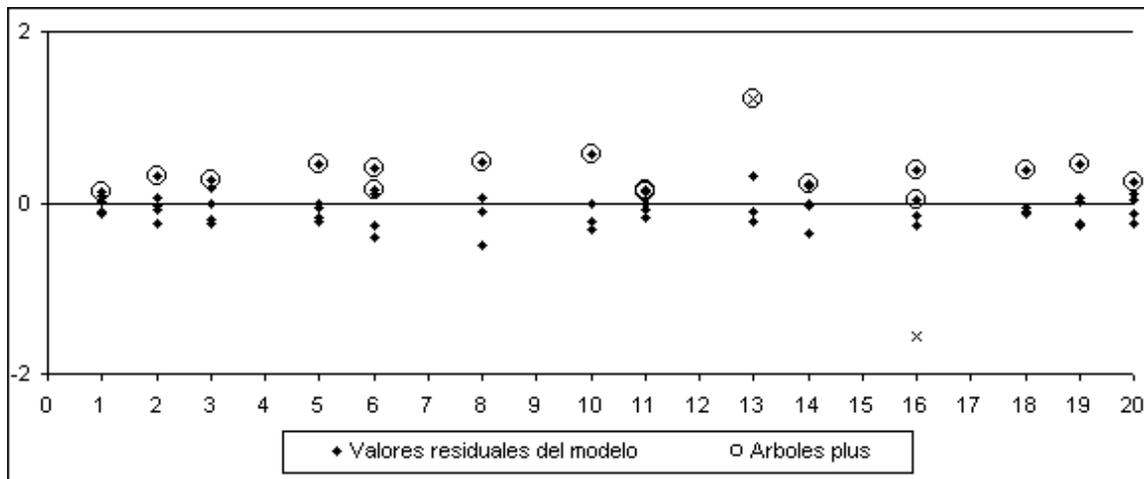


Figura 1. Residuos del modelo de producción individual en 14 rodales (de 1 a 20) como criterio de selección de 1 ó 2 árboles plus por rodal. ( x - 2 árboles excluido de la modelización, por su valor residual excesivo y su gran influencia)