

AJUSTE ÓPTIMO DE LA DISTRIBUCIÓN LOGNORMAL DE TRES PARÁMETROS EN HÍBRIDOS DE POPULUS

O. S. Vallejos-Barra

Profesor de la Facultad de Ciencias Forestales de la Universidad de Talca, 2 Norte 685, Talca, Chile.
Correo-e: ovallejo@utalca.cl. Número telefónico (56) (71) 200445.

RESUMEN

La variable diámetro a la altura del pecho, medido a 1.3 m desde el suelo (DAP) es lognormal si $\log(\text{DAP}-\theta)$ se distribuye normalmente. Para simplificar el ajuste de esta función de densidad de probabilidad (fdp) se supone que el parámetro de localización (θ) es cero en híbridos de *Populus*; sin embargo, en algunos casos esta simplificación no asegura el mejor ajuste de la fdp, incidiendo en la representación de la estructura horizontal de los híbridos. El objetivo de la investigación fue calcular el parámetro θ que proporcione el mejor ajuste de la fdp lognormal del DAP en los híbridos I-488 e I-63/51. Para este cálculo se utilizó un algoritmo de búsqueda binaria, que minimizó el estadístico D_n de la prueba de Kolmogorov-Smirnov. 38 % de las parcelas presentó un parámetro θ distinto de cero, originando una mejoría del valor de probabilidad (P) entre 0.02 y 55.4 %. Se hizo un análisis de varianza no paramétrico de la variable " $\sqrt{n} * D_n$ ", ya que la variable original y múltiples transformaciones no permitieron aceptar los supuestos del análisis paramétrico. El tipo de ajuste presentó diferencias estadísticas altamente significativas con un valor de 17.1 ($P < 0.001$) de la prueba de Kruskal-Wallis. Para los híbridos y edades consideradas no hubo diferencias estadísticas, los híbridos presentaron un valor de 0.1 ($P > 0.800$) y las edades un valor de 11.2 ($P > 0.260$).

PALABRAS CLAVE: parámetro de localización, optimización y Kolmogorov-Smirnov

OPTIMAL ADJUSTMENT OF THE THREE-PARAMETERS LOGNORMAL DISTRIBUTION IN HYBRID OF POPULUS

SUMMARY

The DBH variable will be lognormal if $\log(\text{DBH}-\theta)$ it is Normal distributed. To simplify the adjustment of this probability density function (fdp), the localization parameter (θ) is assumed zero in hybrid of *Populus*; however, in some cases this simplification doesn't guarantee the best adjustment, affected the representation of the horizontal structure. In this research, the parameter that provides the best adjustment in the fdp lognormal of the DBH in the hybrid I-488 and I-63/51 was calculated. The algorithm of binary search was used which minimizes the statistical D_n of the Kolmogorov-Smirnov test. 38 % of the presented a parameter different from zero, originating a P-Value improvement between 0.02 and 55.4 %. A non parametric variance analysis of the variable " $\sqrt{n} * D_n$ " was made, since this variable and multiple transformations did not allow to accept parametric analysis assumptions. The adjustment type presented highly significant statistical differences with a value of 17.1 ($P\text{-Value} < 0.001$) of the Kruskal-Wallis test. Both hybrids and ages considered did not have statistical differences, hybrids presented a value of 0.1 ($P\text{-Value} > 0.800$) and ages a value of 11.2 ($P\text{-Value} > 0.260$).

KEY WORDS: localization parameter, optimization and Kolmogorov-Smirnov.

INTRODUCCIÓN

En la literatura se encuentra diversas funciones de densidad de probabilidad (fdp) para la representación de la distribución de los diámetros de los árboles a la altura del pecho (estructura horizontal). Una fdp será propuesta para

una especie o grupo de especies en una zona geográfica en particular, en la medida que supere a las restantes funciones en la capacidad de modelar la estructura horizontal de los datos reales. La prueba de bondad de ajuste de Kolmogorov-Smirnov (KS) se utiliza para determinar la capacidad de modelación de las fdp, cuyo estadístico D_n mide la máxima

distancia entre la probabilidad acumulada observada y estimada (expresión [1]).

$$D_n = \max \left(\max_i \left(\frac{i}{n} - \hat{F}(DAP) \right); \max_i \left(\hat{F}(DAP) - \frac{i-1}{n} \right) \right) \quad [1]$$

Donde i es el i -ésimo árbol, n es el total de árboles de la parcela y $\hat{F}(DAP)$ es la probabilidad acumulada estimada considerando la fdp en análisis. Esta prueba es muy sensible a cualquier diferencia del valor central y de la dispersión.

Las funciones Weibull 3P y S_B de Johnson tienen en común que incorporan un parámetro de localización (a y ε respectivamente) y que la mayoría de las veces se debe suponer su valor. Para la distribución Weibull 3P se ha utilizado el valor $0.50 * DAP$ mínimo (Cao, 2004; Liu *et al.*, 2004). Otras proporciones empleadas han sido: 0.00, 0.33, 0.66 y 1.00 (Liu *et al.*, 2004). También se ha propuesto que este parámetro se obtenga restando una constante al DAP mínimo, cuyos valores han sido: 0.5; 1.0, 1.5 y 2.0 cm (Zhang *et al.*, 2003). En la función S_B de Johnson se debe seleccionar una de las alternativas siguientes i) suponer que $\varepsilon = 0$ (Kudus *et al.*, 1999), ii) calcular ε con el método de Knoebel y Burkhart donde $\varepsilon = DAP$ mínimo - 1.3 o iii) suponer que ε es una proporción del DAP mínimo (Scolforo *et al.*, 2003).

Con el objetivo de evitar suponer el valor del parámetro de localización en el ajuste de la fdp Weibull 3P, Vallejos *et al.* (2004) ajustó estas funciones de manera óptima, al minimizar el estadístico D_n de la prueba de KS. A través de un proceso de búsqueda binaria utilizando el complemento "Solver.xls" de Microsoft Excel®, se estiman los parámetros de cada distribución que minimizan el estadístico D_n .

La presente investigación es una ampliación de esta propuesta metodológica considerando que la fdp lognormal (expresión [2]) ha sido recomendada para el híbrido de álamo I-63/51 en la zona de estudio de esta investigación y que se supone que el parámetro θ tiene un valor de cero. El objetivo es evidenciar si el parámetro θ difiere de cero y si

esto es así, si existen diferencias estadísticas en el tipo de ajuste (ajuste tradicional, con $\theta = 0$ y ajuste óptimo, con $0 < \theta < DAP$ mínimo).

$$f(DAP) = \frac{1}{(DAP - \theta) \sigma \sqrt{2\pi}} \exp \left(-\frac{(\log(DAP - \theta) - \mu)^2}{2\sigma^2} \right) \quad [2]$$

Donde θ es el parámetro de localización ($0 \leq \theta < DAP$ mínimo), μ es el parámetro de escala y σ es el parámetro de forma, debiendo ser ambos positivos. Según Johnson *et al.* (1995) los estimadores de los parámetros de escala y de forma se obtienen utilizando las expresiones [3] y [4] respectivamente.

$$\hat{\mu} = \frac{\sum_{i=1}^n \log(DAP_i - \theta)}{n} \quad [3]$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (\log(DAP_i - \theta) - \hat{\mu})^2}{n}} \quad [4]$$

MATERIALES Y MÉTODOS

Los datos provienen de 243 parcelas del híbrido I-488 y de 190 parcelas del híbrido I-63/51, pertenecientes a la Compañía Agrícola y Forestal El Álamo Limitada (36° 05' latitud sur y 72° 47' longitud este). En el Cuadro 1, se presentan los datos que caracterizan a ambos híbridos, observándose que el híbrido I-63/51 presenta árboles de mayores dimensiones en todas las edades, confirmando lo indicado por Loewe *et al.* (1997) con relación a la diferencia de tamaño entre los híbridos.

Para el ajuste de las parcelas se utilizó el complemento Optilogn.xls, desarrollado por el autor de esta publicación

CUADRO 1. Descripción de las parcelas consideradas en esta investigación.

| Híbrido | Variable | Edad (años) | | | | | | | | | | Total |
|---------|-------------------|-------------|------|------|------|------|------|------|------|------|------|-------|
| | | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | |
| I-488 | Núm. parcelas | 9 | 17 | 28 | 31 | 34 | 39 | 39 | 28 | 14 | 4 | 243 |
| | DAP mínimo (cm) | 7.4 | 9.5 | 12.1 | 13.2 | 14.2 | 15.4 | 16.0 | 20.4 | 20.8 | 24.2 | |
| | DAP promedio (cm) | 14.9 | 18.1 | 20.6 | 22.3 | 24.0 | 25.2 | 26.3 | 27.7 | 29.5 | 29.4 | |
| | DAP máximo (cm) | 20.6 | 24.3 | 26.3 | 29.8 | 29.3 | 31.2 | 32.7 | 34.1 | 35.4 | 35.7 | |
| I-63/51 | Núm. parcelas | 6 | 12 | 15 | 23 | 23 | 30 | 32 | 29 | 17 | 3 | 190 |
| | DAP mínimo (cm) | 14.4 | 15.0 | 16.9 | 18.8 | 20.0 | 21.9 | 23.3 | 24.6 | 26.0 | 31.0 | |
| | DAP promedio (cm) | 19.3 | 23.6 | 26.0 | 27.5 | 29.3 | 30.5 | 31.7 | 33.0 | 33.7 | 34.1 | |
| | DAP máximo (cm) | 23.2 | 28.2 | 30.0 | 32.5 | 34.4 | 35.9 | 37.3 | 38.9 | 39.1 | 37.2 | |

(disponible en www.cienciasforestales.cl donde se debe seleccionar "UTILIDADES" y posteriormente "Publicaciones de archivos y textos de interés para la comunidad estudiantil").

Se realizó el ajuste tradicional ($\theta = 0$), obteniendo los estimadores de la fdp (expresiones [3] y [4]) y calculando el estadístico D_n . Para el ajuste óptimo Optilogn.xla interactúa con el complemento Solver.xla, disponible en Microsoft Excel®, minimizando el estadístico D_n de la prueba de KS (expresión [5]) considerando las restricciones indicadas en la expresión [6].

Función objetivo:

$$\text{Mínimo} \left\{ D_n = \max \left(\max_i \left(\frac{i}{n} - F(DAP) \right); \max_i \left(F(DAP) - \frac{i-1}{n} \right) \right) \right\} \quad [5]$$

Sujeto a:

$$\begin{aligned} \theta &\geq 0 \\ \theta &< \text{DAP mínimo} && [6] \\ \hat{\mu} &> 0 \\ &> 0 \end{aligned}$$

Donde $F(DAP) = \Phi(\log(DAP - \theta))$ y Φ es la función de distribución normal acumulativa.

Se realizó un análisis de varianza de ambos ajustes ($\theta = 0$ y θ óptimo), considerando como variable explicativa $\sqrt{n} * D_n$, ya que la cantidad de árboles por parcela no fue constante. Los supuestos del análisis de varianza no se cumplieron, de modo que se utilizó la prueba de Kruskal-Wallis, que corresponde a un análisis de varianza no paramétrico.

RESULTADOS

Se encontró que 97 parcelas del híbrido I-488 presentaron un parámetro θ distinto de cero, mientras que para el híbrido I-63/51 fueron 66 parcelas, representando 40 y 35 por ciento de la cantidad total de parcelas, respectivamente.

En la Figura 1 se presenta la frecuencia de parcelas de cada decil del parámetro θ relativo (división del parámetro θ por el DAP mínimo de cada parcela). En términos generales se observa en ambos híbridos mayor frecuencia a medida que aumenta el decil, con una concentración superior al 75 % sobre el séptimo decil. Claramente en estos casos suponer un θ igual a cero no proporciona el mejor ajuste.

En el Cuadro 2 se aprecia el resultado del análisis de varianza no paramétrico (prueba de Kruskal-Wallis). La única

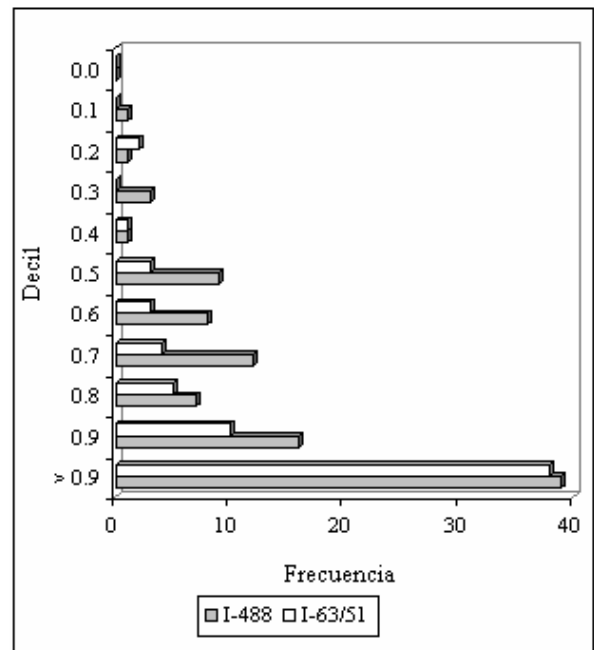


FIGURA 1. Frecuencia del parámetro θ relativo por decil según híbrido.

fuerza de variación que presentó diferencias estadísticas fue el tipo de ajuste, siendo altamente significativas. Como era de esperar, al considerar sólo las parcelas cuyo parámetro θ fue distinto de cero, el valor de la prueba superó al obtenido considerando la totalidad de las parcelas. En el caso de los híbridos y de la edad, este aumento no permitió detectar diferencias estadísticas.

CUADRO 2. Prueba de Kruskal-Wallis de $\sqrt{n} * D_n$ según fuente de variación y cantidad de parcelas.

| Fuente de variación | Todas las parcelas | Parcelas con $\theta \neq 0$ |
|---------------------|--------------------|------------------------------|
| Tipo de ajuste | 17.1*** | 56.7*** |
| Híbridos | 0.1 ^{NS} | 7.1 ^{NS} |
| Edad | 11.2 ^{NS} | 15.3 ^{NS} |

***: Diferencias estadísticas altamente significativas ($P < 0.001$).
^{NS}: Diferencias estadísticas no significativas ($P > 0.05$).

En el Cuadro 3, se presenta la comparación del tipo de ajuste. El ajuste óptimo considerando todas las parcelas o sólo aquellas con un parámetro θ distinto de cero presentó el menor promedio y por lo tanto modela de mejor forma la estructura horizontal que el ajuste tradicional.

CUADRO 3. Comparación de $\sqrt{n} * D_n$ promedio según tipo de ajuste

| Tipo de ajuste | Todas las parcelas | Parcelas con $\theta \neq 0$ |
|------------------------------|--------------------|------------------------------|
| Óptimo ($\theta \neq 0$) | 0.61 a' | 0.48 a |
| Tradicional ($\theta = 0$) | 0.65 b | 0.59 b |

*Letras distintas indican diferencias estadísticas altamente significativas, según Kruskal-Wallis ($P < 0.001$).

CONCLUSIONES

Se comprobó en los híbridos de *Populus* I-488 e I-63/51 que el parámetro θ difiere del valor cero en 38 % de los casos. En términos generales los valores del parámetro θ se encuentran arriba de la mitad del DAP mínimo con una alta concentración alrededor del DAP mínimo.

El ajuste óptimo de la función de densidad lognormal presentó diferencias estadísticas altamente significativas con relación al ajuste tradicional, de modo que se recomienda su utilización en posteriores investigaciones.

LITERATURA CITADA

- CAO, Q. V. 2004. Predicting parameters of a Weibull function for modeling diameter distribution. *For. Sci.* 50(5): 682 – 685.
- JOHNSON, N. L.; KOTZ, S.; BALAKRISNAN, N. 1995. *Continuous Univariate Distributions*. Volume 1. Second Edition. John Wiley & Sons, Inc. 756p.
- KUDUS, K. A.; AHMAD, M. I.; LAPONGAN, J. 1999. Nonlinear regression approach to estimating Johnson S_B parameters for diameter data. *Can. J. For. Res.* 29: 310 - 314
- LIU, CH.; ZHANG, S. Y.; LEI, Y.; NEWTON, P.; ZHANG, L. 2004. Evaluation of three methods for predicting diameter distributions of black spruce (*Picea mariana*) plantations in central Canada. *Can. J. For. Res.* 34: 2424- 2432.
- LOEWE, V.; TORAL, M.; FERNÁNDEZ, M.; PINEDA, G.; LÓPEZ, C.; URQUIETA, E. 1997. Potencialidad de especies y sitios para una diversificación silvícola nacional: Monografía de Álamo *Populus* spp. INFOR – CONAF. 143p.
- SCOLFORO, J. R.; TABAI, F. C. V.; MACEDO, R. L. G.; ACERBI, F. W.; ASSIS, A. L. 2003. S_B distribution's accuracy to represent the diameter distribution of *Pinus taeda*, through five fitting methods. *For. Ecol. Manage.* 175: 489 – 496.
- VALLEJOS, O. S.; SANQUETTA, C. R.; ARCE, J. E.; MACHADO, S. do A.; DALLA, A. P. 2004. Proposta metodológica para o ajuste ótimo da distribuição diamétrica Weibull 3P. *Floresta* 34 (3): 387-393.
- ZHANG, L.; PACKARD, K.; LIU, Ch. 2003. A comparison of estimation methods for fitting Weibull and Johnson's S_B distributions to mixed spruce-fir stands in northeastern North America. *Can. J. For. Res.* 33: 1340 – 1347.