

Weibull_fit.xls: Programa para el Ajuste Óptimo de la Función de Densidad de Probabilidad Weibull de Tres Parámetros

Oscar S. Vallejos y Darío M. Aedo

Universidad de Talca, Facultad de Ciencias Forestales, Departamento de Producción Forestal, 2 Norte 685, Talca-Chile (e-mail ovallejo@utalca.cl)

Resumen

En este trabajo se da a conocer el programa de dominio público denominado Weibull_fit.xls que amplió la propuesta original de los autores. La nueva propuesta considera como función objetivo, además de la prueba de Kolmogorov-Smirnov, las restantes pruebas de bondad de ajuste y el índice de error. Se introdujo esta modificación por la incorporación del índice de error que trabaja con clases diamétricas y porque matemáticamente es posible obtener resultados negativos del parámetro de localización. Como ejemplo se modelaron los diámetros de árboles de una parcela de *Pinus radiata* (D. Don) con el método de los momentos, considerando cada indicador de error. Se concluye que el programa Weibull_fit.xls es una herramienta computacional que permite estimar los parámetros de la función Weibull 3P que minimizan los estadígrafos de las pruebas de bondad de ajuste o del índice de error.

Palabras clave: Weibull_fit.xls, ajuste óptimo, función densidad, Weibull 3P, diámetro de árboles

Weibull_fit.xls: Software for Optimum Adjustment of the Probability Density Function Weibull 3P

Abstract

The probability density function Weibull 3P is broadly used in the literature to represent continuous data. Most of the software that adjusts this distribution utilizes a specific parameter estimation method that calculates values for only two parameters, assuming the value for the third one. On the other hand, the generated public domain program calculates the three parameters that minimize the chosen error indicator (six in total) for the selected parameter estimation method (four in total). The potential error indicators to be used are associated with the statistic scores of five goodness-of-fit and the index of error. As an example the tree diameters data of a *Pinus radiata* (D. Don) plot was modeled by the moment method, considering each of the error indicators. In each adjustment was verified that the minimum error was reached for the calculated parameters.

Keywords: Weibull_fit.xls, optimum fitting, density function, Weibull 3P, tree diameter

INTRODUCCIÓN

La función de densidad de probabilidad (fdp) Weibull (1) fue propuesta inicialmente por Fisher y Tippett en el año 1928. Esta función adoptó el apellido del físico sueco Waloddi Weibull, quien la difundió a partir del año 1939 en estudios de resistencia de materiales (Johnson *et al.*, 1994), siendo ampliamente usada en distintas áreas del conocimiento para modelar datos continuos. Su amplia utilización radica en la representación de diversas formas, de este modo si el parámetro de forma es menor que uno la forma será decreciente, si es uno la forma será exponencial, si es dos la forma será Rayleigh, si es tres coma seis la forma será normal y si es mayor que tres coma seis la forma será normal pero con asimetría negativa (Nord-Larsen y Cao, 2006). Además posee integral definida (2) permitiendo el cálculo confiable de probabilidades (Borders *et al.*, 1987). Finalmente, Liu *et al.* (2004) y Mehtätalo (2004) resaltan la relativa facilidad para estimar sus parámetros.

$$f(x) = \frac{c}{b} \left(\frac{x-a}{b} \right)^{c-1} * \exp \left(- \left(\frac{x-a}{b} \right)^c \right) \quad (1)$$

$$F(x) = 1 - \exp \left(- \left(\frac{x-a}{b} \right)^c \right) \quad (2)$$

Donde a , b y c son los parámetros de localización, escala y forma respectivamente y x es una variable continua.

En el sector forestal los precursores de la utilización de esta fdp fueron Bailey y Dell en el año 1973. Su uso en este sector se ha concentrado en modelar la estructura de los rodales, pudiendo conocer la frecuencia relativa y acumulada de los diversos tamaños de los árboles (Prodan *et al.*, 1997). El tamaño de los árboles, es decir la variable x continua de la función, corresponde mayoritariamente al diámetro a la altura del pecho (DAP) de los árboles medidos en cada unidad muestral. Diversos autores han comprobado que la fdp Weibull caracteriza de manera confiable los rodales considerados en sus investigaciones (Jiang y Brooks, 2009; Breidenbach *et al.*, 2008; Castedo-Dorado *et al.*, 2007; Palahi *et al.*, 2007; Vallejos, 2003)

Si bien es cierto que conocer los parámetros de la fdp en un momento dado es importante, aún más relevante es poder proyectar el tamaño de los árboles en el tiempo (Prodan *et al.*, 1997). Una de las técnicas usadas es el método de predicción de parámetros, el cual consiste en relacionar los parámetros de la fdp con variables que caracterizan el rodal. Diversos autores han utilizado este método con la fdp Weibull caracterizando de manera confiable los rodales considerados en sus investigaciones (Nord-Larsen y Cao, 2006 y Palahi *et al.*, 2006)

Para estimar los parámetros de esta fdp se han descrito diversos métodos de estimación, algunos de ellos tienen larga data de publicación, ya que son procedimientos matemáticos validados hace mucho tiempo. Se puede mencionar entre otros (Tabla 1): el método híbrido descrito por Cao (2004), el método de máxima verosimilitud descrito por Johnson *et al.* (1994), el método de los momentos descrito por Cao (2004) y el método de los percentiles descrito por Scolforo (1998).

En los métodos de estimación de la Tabla 1 es necesario asumir un valor para el parámetro de localización. Este parámetro ha sido considerado como un porcentaje del DAP mínimo; frecuentemente se ha utilizado el valor 0.50 (Frazier, 1981, citado por Cao, 2004; Hawkins *et al.*, 1988 y Liu *et al.*, 2004). Otros porcentajes empleados han sido: 0.00, 0.33, 0.66 y 1.00 (Liu *et al.*, 2004). Nepal y Somers (1992) asumen que el parámetro de localización es el límite inferior de la primera clase diamétrica. Finalmente 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).

Tabla 1: Métodos de estimación de parámetros de la fdp Weibull 3P

Método y autor	Expresión
Híbrido (Cao, 2004)	$b = \frac{x_{93} - a}{(-\ln(1-0,93))^{1/c}}$ $a^2 - x_q^2 + 2*a*\frac{(x_{93} - a)*\Gamma_1}{(-\ln(1-0,93))^{1/c}} + \frac{(x_{93} - a)^2*\Gamma_2}{(-\ln(1-0,93))^{2/c}} = 0$
Máxima verosimilitud (Johnson <i>et al.</i> 1994)	$b = b'^{1/c}; b' = \frac{\sum_{i=1}^n (x_i - a)^c}{n}$ $\frac{1}{c} = \frac{\sum_{i=1}^n (x_i - a)^c * \ln(x_i - a)}{\sum_{i=1}^n (x_i - a)^c} - \frac{\sum_{i=1}^n (x_i - a)}{n}$ $(c-1) * \sum_{i=1}^n \frac{1}{(x_i - a)} = \frac{c}{b^c} * \sum_{i=1}^n (x_i - a)^{c-1}$
Momentos (Cao, 2004)	$b = \frac{(\bar{x} - a)}{\Gamma_1}$ $x_q^2 + a^2 - 2*a*\bar{x} - b^2*\Gamma_2 = 0$
Percentiles Scolforo (1998)	$c = \frac{\sum_{i=1}^n xx_j^2 - \frac{(\sum_{i=1}^n xx_j)^2}{n}}{\sum_{i=1}^n xx_j * y_j - \frac{\sum_{i=1}^n xx_j * \sum_{i=1}^n y_j}{n}}$ $b = \exp\left(\frac{\sum_{i=1}^n y_j}{n} - \frac{1}{c} * \frac{\sum_{i=1}^n xx_j}{n}\right)$

Donde x es la variable continua en estudio, n es la cantidad total de observaciones, x_i es la observación i -ésima de la variable x , x_q es la media cuadrática de la variable x , \bar{x} es la media aritmética de la variable x , Γ_i es la función gamma de $\Gamma_{(1 + i/c)}$, x_{93} es el valor de la variable x al percentil 93, xx_j es $\ln(-\ln(1 - j / 10))$ e y_j es $\ln(x_j)$, siendo $j = 10, 20, \dots, 80, 90$ percentiles.

Vallejos (2003), definió un método numérico que estima incluso el parámetro de localización, evitando el uso de un supuesto que limitaba las capacidades de ajuste de la función. Vallejos *et al.* (2004) indican que mediante un proceso iterativo de búsqueda binaria usando el algoritmo Newton-Raphson, se calculan los tres parámetros (4), dependiendo del método de estimación seleccionado (Tabla 1), que minimizan el estadístico D_n de la prueba de bondad de ajuste de Kolmogorov-Smirnov (3). También, Chen (2006) utilizó este algoritmo para el ajuste de la fdp Lognormal de tres parámetros.

$$\text{Función objetivo: Minimizar } D_n = \max\left(\max\left(\frac{i}{n} - F(DAP_i)\right); \max\left(F(DAP_i) - \frac{i-1}{n}\right)\right) \quad (3)$$

Sujeto a:

$$\begin{array}{rclcl} 0 & \leq & a & \leq & \text{DAP mínimo} \\ & & b & > & 0 \\ & & c & > & 0 \end{array} \quad (4)$$

Donde i varía entre 1 a n (contador desde el menor al mayor valor del DAP) y $F(DAP_i)$ que ya fue definido en la ecuación 2.

La presente publicación da a conocer el programa de dominio público denominado Weibull_fit.xls que amplió la propuesta original de Vallejos (2003) y Vallejos *et al.* (2004), ya que considera como función objetivo (Tabla 2), además de la prueba de Kolmogorov-Smirnov las restantes pruebas de bondad de ajuste (Anderson-Darling, Kuiper, Cramer-Von Mises y Watson) y el índice de error (Reynolds *et al.* 1988). Además se modificó la restricción asociada al parámetro de localización, ya que este podrá ser menor o igual que el límite inferior de la primera clase diamétrica. Se introdujo esta modificación por la incorporación del índice de error que trabaja con clases diamétricas y porque matemáticamente es posible obtener resultados negativos del parámetro de localización.

Tabla 2: Descripción de la función objetivo

Función objetivo	Expresión
Pruebas de bondad de ajuste	
Kolmogorov-Smirnov	$D_n = \max\left(\max\left(\frac{i}{n} - F(DAP_i)\right); \max\left(F(DAP_i) - \frac{i-1}{n}\right)\right)$
Anderson-Darling	$A^2 = -n - \frac{\sum_{i=1}^n (2 * i - 1) * \ln(F(DAP_i)) + (2 * n + 1 - 2 * i) * \ln(1 - F(DAP_i))}{n}$
Kuiper	$V = \max\left(\frac{i}{n} - F(DAP_i)\right) + \max\left(F(DAP_i) - \frac{i-1}{n}\right)$
Cramer-Von Mises	$W^2 = \sum_{i=1}^n \left(F(DAP_i) - \frac{2 * i - 1}{2 * n}\right)^2 + \frac{1}{12 * n}$
Watson	$U^2 = W^2 - n * (\overline{F(DAP)} - 0,5)^2$
Índice de error	$IE = \sum_{j=1}^k n_j - \hat{n}_j $

Donde $\overline{F(DAP)}$ es el promedio de los $F(DAP_i)$, n_j y \hat{n}_j son el número de elementos medidos y estimados respectivamente en la clase j y k es el total de clases (como la información continua puede ser muy extensa, la variable DAP se discretiza generando clases diamétricas de amplitud predefinida)

MÉTODOLÓGIA

Weibull_fit.xls fue programado en Microsoft Visual Basic 6.3, que una vez instalado se presenta como un menú desplegable en Microsoft Excel (Figura 1). El programa puede solicitarse gratuitamente llenando el formulario de contacto disponible en: <http://www.cienciasforestales.cl/contacto/demo.php>.

En la Figura 2 se presenta la secuencia de programación de Weibull_fit.xls, que entra en ejecución al seleccionar el módulo Calculo (Figura 1).

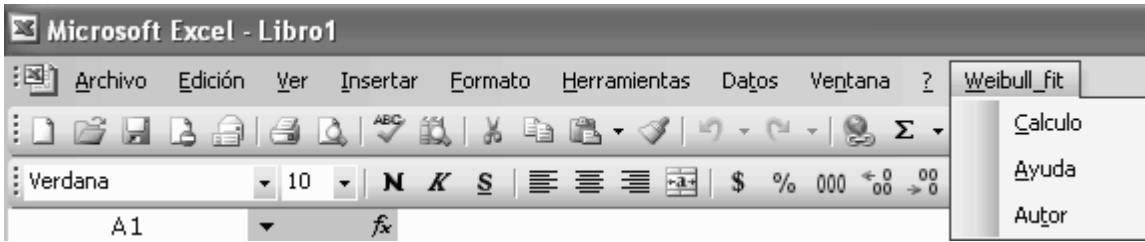


Fig. 1: Menú desplegable del programa Weibull_fit.xls

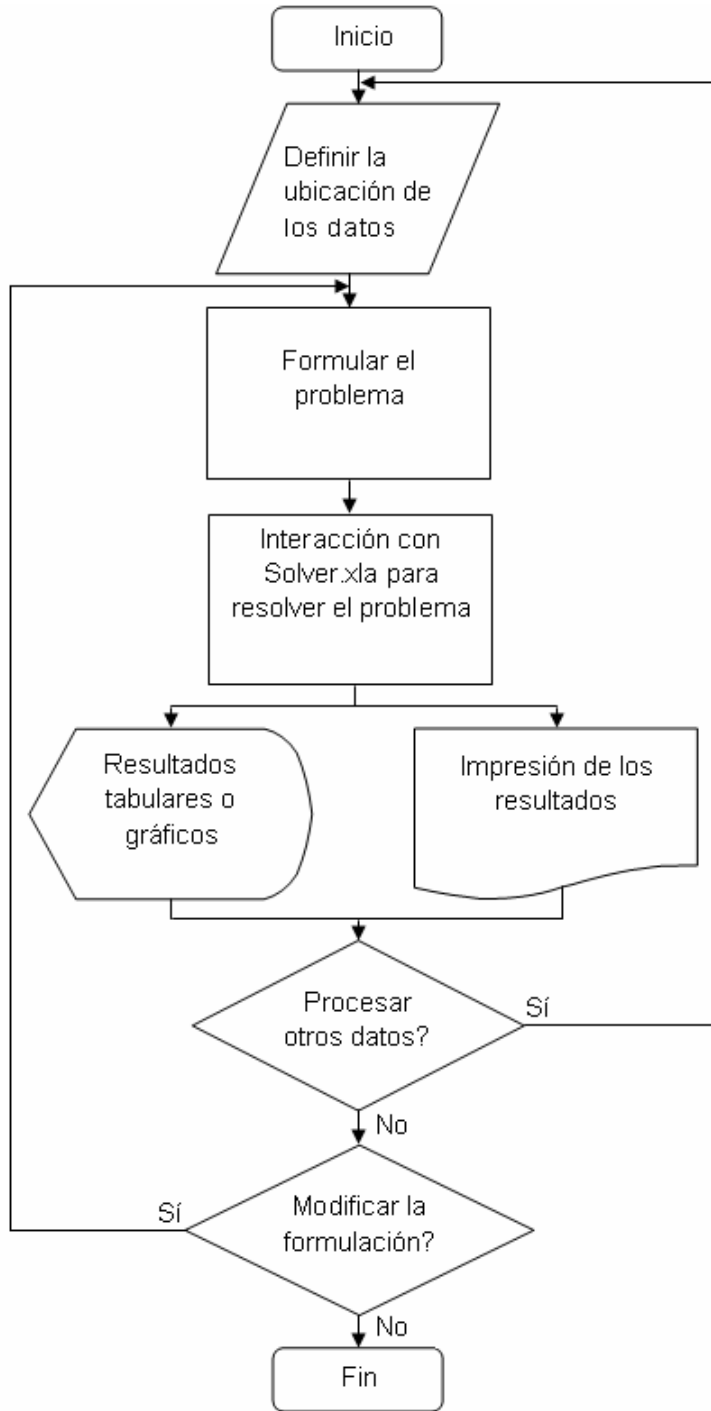


Fig. 2: Diagrama de flujo de la programación de Weibull_fit.xls

En primer lugar se debe indicar la ubicación de la muestra aleatoria de datos en estudio. En segundo lugar se formula el problema, para esto se selecciona una función objetivo, especificando qué prueba

de bondad de ajuste o índice de error se utilizará (Tabla 2) y se definen las restricciones asociadas al método escogido para la estimación de parámetros (Tabla 1). En tercer lugar, el programa interactúa con Solver.xls que lleva a cabo la minimización mediante el algoritmo de Newton Raphson. Una vez que se logra un mínimo de la función objetivo, se presentan los resultados en forma tabular y se despliega un formulario que presenta las opciones del usuario (Figura 3). Es posible observar o imprimir la información tabular o gráfica, modificar la formulación del problema, cambiando la función objetivo o las restricciones, procesar otra parcela o bien salir del programa.



Fig. 3: Opciones posteriores a la minimización.

Como ejemplo se hizo la minimización de todas las pruebas de bondad de ajuste y del índice de error considerando el método de los momentos para estimar los parámetros. Los datos provienen de la parcela descrita en la página 198 de Prodan *et al.* (1997) que corresponde a una parcela de *Pinus radiata* (D. Don) de 400 m² con 41 árboles de 20 años de edad.

RESULTADOS Y DISCUSION

Se debe indicar la ubicación de los datos de la parcela (Figura 4). Estos deben estar dispuestos en una columna.



Fig. 4: Definición de los datos de la parcela.

Para formular el problema se debe escoger la función objetivo y las restricciones (Figura 5). Cuando la función objetivo corresponde a la minimización del índice de error, además se solicitará la amplitud de las clases diamétricas, donde se puede escoger desde 1 a 10 cm. De igual modo, si el método de los percentiles es escogido, se tendrá que indicar el valor de éstos, pudiendo escoger entre 1 y 100. Los trabajos de Dubey (1967) y Bailey y Dell (1973) consideraron dos percentiles. Scolforo (1998) aumentó a nueve el máximo de percentiles posibles de ser definidos pero, Weibull_fit.xls permite una completa libertad en la definición de ellos.

Los resultados se presentan en forma tabular (Figura 6), donde se indica la función objetivo, las restricciones (métodos de estimación de parámetros), los parámetros de la fdp Weibull 3P, los estadígrafos de cada prueba de bondad de ajuste y el valor porcentual del índice de error, el cual se obtiene al dividir el índice de error (Tabla 2) por la cantidad de elementos medidos (n). Además, se presentan dos resultados donde se asume el parámetro de localización igual a cero o igual a la mitad del diámetro mínimo ($a = 0$ y $a = \text{DAP mínimo}/2$). Estos valores se incluyeron ya que son los más recomendados en la literatura y sirven para establecer una comparación con los resultados minimizados.



Fig. 5: Formulación del problema.

Función objetivo	Restricciones	a	b	c	A^2	W^2	IE %	Dn	Probabilidad	V	U^2
Anderson Darling (A^2)	Momentos	5,342170681	19,55340337	2,810899325	0,2111	0,0330	57,7	0,0732	0,9800	0,1352	0,0330
Cramer-Von Mises (W^2)		6,964730453	17,79426712	2,519994327	0,2251	0,0318	58,3	0,0648	0,9960	0,1288	0,0316
Indice de error (IE)		8,9	15,64589764	2,177837005	0,3294	0,0344	36,5	0,0697	0,9874	0,1303	0,0329
Kolmogorov-Smirnov (Dn)		7,086684854	17,66067662	2,498264532	0,2278	0,0318	36,8	0,0641	0,9960	0,1282	0,0316
Kuiper (V)		8,284249708	16,33657019	2,286045156	0,2763	0,0329	36,7	0,0660	0,9945	0,1228	0,0320
Watson (U^2)		7,337856827	17,38485576	2,453576425	0,2342	0,0319	36,8	0,0644	0,9960	0,1270	0,0315
Sin minimizar (a=0)		0	25,18378128	3,786000441	0,2668	0,0450	37,1	0,0927	0,8772	0,1507	0,0432
Sin minimizar (a=Dmin/2)		5	19,92057889	2,872606626	0,2115	0,0335	36,9	0,0748	0,9753	0,1364	0,0335

Fig. 6: Resultado tabular del ejemplo.

Según la conveniencia del usuario es posible obtener el resultado de un ajuste en particular hasta las 32 posibilidades que entrega el programa [4 métodos de estimación de parámetros * (6 indicadores de error ha minimizar) + 2 alternativas que asumen un valor para el parámetro de localización)].

Se observa que las estimaciones de los parámetros sin minimizar, generan estadígrafos superiores a los obtenidos al aplicar minimización. También se comprueba que el valor mínimo de cada estadígrafo se alcanza cuando dicho estadígrafo es considerado como la función objetivo a minimizar. Esto muestra que Weibull_fit.xls emplea un modelo optimizante para el ajuste de los parámetros de la función Weibull 3P.

En la Figura 7 se presenta el resultado gráfico de los ajustes. Al ejecutar el programa las líneas asociadas a cada ajuste aparecen en color, de modo que es posible distinguir entre ellas. Se observa que pese a lograr resultados estadísticos significativos, probabilidad asociada a la prueba Kolmogorov-Smirnov varía entre 87.7% y 99.6%, la estimación del número de árboles por clase diamétrica no logra una adecuada representación, ya que el índice de error porcentual fluctúa entre un 36.5% y un 58.3%.

Las limitaciones del programa Weibull_fit.xls están asociadas a la sincronía con el complemento Solver.xls de Microsoft Excel®. En la versión 2000 y 2003 la sincronía es perfecta, lo cual no ocurre en la versión 2002 y 2007 donde el programa entrega un error de aplicación y no puede ser ejecutado. En futuras versiones del programa se permitirá la exportación de la tabla de resultados.

Si un usuario requiere desarrollar un método de predicción de parámetros, con la finalidad de proyectar el tamaño de los árboles en el tiempo, encontrará en Weibull_fit.xls una herramienta para la obtención de los parámetros. La relación entre los parámetros y las variables que caracterizan un rodal tendrá que desarrollarse con cualquier programa estadístico comercial.

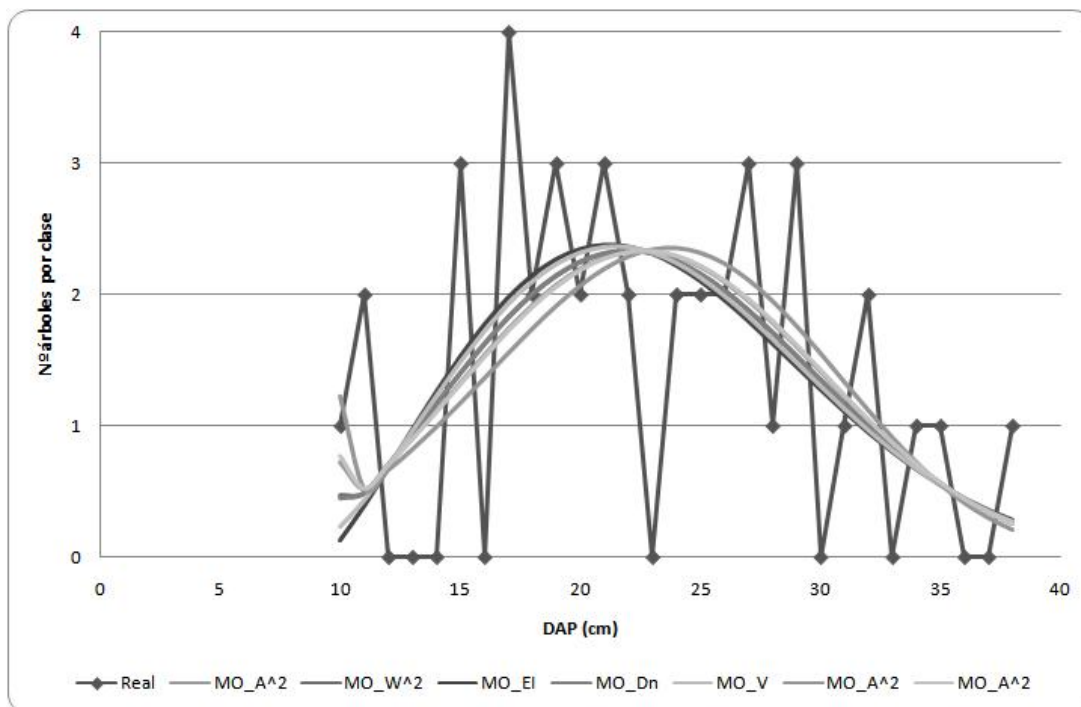


Fig. 7: Resultados gráfico del ejemplo.

CONCLUSIONES

El programa Weibull_fit.xls es una herramienta computacional que permite estimar los parámetros de la función Weibull 3P que minimizan los estadígrafos de las pruebas de bondad de ajuste o del índice de error.

El programa presenta en una sola tabla todos los resultados obtenidos, de modo que es posible establecer comparaciones entre los métodos de ajuste (restricciones), o entre las pruebas de bondad de ajuste e índice de error (funciones objetivos alternativas).

REFERENCIAS

Bailey, R. y T. Dell; *Quantifying diameter distributions with the Weibull function*, Forest Science: 19(2); 97 – 104 (1973).

Borders B. y otros tres autores; *Percentile-based distributions characterize forest stand tables*, Forest Science: 33(2); 570 – 576 (1987).

Breidenbach, J., C. Glaser y M. Schmidt; *Estimation of diameter distributions by means of airborne laser scanner data*, Canadian Journal Of Forest Research, 38(6); 1611 – 1620 (2008).

Cao Q.; *Predicting parameters of a Weibull function for modeling diameter distribution*, Forest Science: 50(5); 682 – 685 (2004).

Castedo-Dorado, F., U. Dieguez-Aranda y JG Alvarez-Gonzalez; *A growth model for Pinus radiata D. Don stands in north-western Spain*, Annals of Forest Science: 64(4); 453 – 465 (2007).

Chen C.; *Test of fit the three-parameter lognormal distribution*, Computational Statistics & Data Analysis: 50; 1418 – 1440 (2006).

Dubey S.; *Some percentile estimators for Weibull parameters*, Technometrics: 9; 119-129 (1967).

Hawkins, K. y otros tres autores; *Using the Weibull distribution to model harvesting machine productivity*, Forest Products Journal: 38(4); 59 – 65 (1988).

- Jiang, L. y J. Brooks; *Predicting Diameter Distributions for Young Longleaf Pine Plantations in Southwest Georgia*, Southern Journal of Applied Forestry: 33(1); 25 – 28 (2009).
- Johnson, N., S. Kotz y N. Balakrishnan; *Weibull distributions, Continuous univariate distributions*, Second Edition, John Wiley & Sons, Inc., volume 1, pp 628 – 722, New York, USA (1994).
- Liu, Ch. y otros cuatro autores; *Evaluation of three methods for predicting diameter distributions of black spruce (Picea mariana) plantations in central*, Canadian Journal of Forest Research: 34; 2424-2432 (2004).
- Mehtätalo L.; *An algorithm for ensuring compatibility between estimates percentiles of diameter distribution and measured stand variables*, Forest Science: 50(1); 20 – 32 (2004).
- Nepal, S. y G. Somers; *A generalized approach to stand table projection*, Forest Science: 38(1); 120 –133 (1992).
- Nord-Larsen, T. y Q. Cao; *A diameter distribution model for even-aged beech in Denmark*, Forest Ecology and Management: 231(1-3); 218 – 225 (2006)
- Palahi, M., T. Pukkala y A. Trasobares; *Modelling the diameter distribution of Pinus sylvestris, Pinus nigra and Pinus halepensis forest stands in Catalonia using the truncated Weibull function*, Forestry: 79(5); 553 – 562 (2006).
- Palahi, M. y otros tres autores; *Comparison of beta, Johnson's SB, Weibull and truncated Weibull functions for modeling the diameter distribution of forest stands in Catalonia (north-east of Spain)*, European Journal of Forest Research: 126(4); 563 – 571 (2007).
- Prodan, M. y otros tres autores; *Mensura Forestal serie investigación y educación en desarrollo sostenible*, IICA/BMZ/GTZ, pp 198, San José, Costa Rica (1997).
- Reynolds, M. Jr., T. Burk y W. Huang; *Goodness-of-fit test and model selection procedures for diameter distribution models*, Forest Science: 34(2); 373 – 399 (1988).
- Scolforo J.; *Distribuição Weibull, Modelagem do crescimento e da produção de florestas plantadas e nativas*, Universidade Federal de Lavras, Fundação de Apoio ao Ensino, Pesquisa e Extensão, pp 103 - 113 (1998).
- Vallejos O.; *Sistema de simulação do crescimento e produção de Populus spp. no Chile (Sistema SALICA)*, Doctorado en Ciencias Forestales, Universidad Federal de Paraná, Sector de Ciencias Agrarias, Departamento de Ciencias Forestales (2003).
- Vallejos, O. y otros cuatro autores; *Proposta metodológica para o ajuste ótimo da distribuição diamétrica Weibull 3P*, Rev. Floresta, ISSN: 1982-4688 (en línea), 34(3); 387-393; 2004. <http://ojs.c3sl.ufpr.br/ojs2/index.php/floresta/issue/view/268>. Acceso: 5 de Mayo (2007).
- Zhang, L.; K. Packard y Ch. Liu. *A comparison of estimation methods for fitting Weibull and Johnson's SB distributions to mixed spruce-fir stands in northeastern North America*, Canadian Journal of Forest Research: 33; 1340 – 1347 (2003).