

Calibrating Predicted Tree Diameter Distributions in Catalonia, Spain

Marc Palahí, Timo Pukkala and Antoni Trasobares

Palahí, M., Pukkala, T. & Trasobares, A. 2006. Calibrating predicted tree diameter distributions in Catalonia, Spain. *Silva Fennica* 40(3): 487–500.

Several probability density functions have been used in describing the diameter distributions of forest stands. In a case where both the stand basal area and number of stems per hectare are assessed, the fitted or predicted distribution is scaled using only one of these variables, with the result that the distribution often gives incorrect values for the other variable. Using a distribution that provides incorrect values for known characteristics means wasting information. Calibrating the distribution so that it is compatible with the additional information on stand characteristics is a way to avoid such wasting. This study examined the effect of calibration on the accuracy of the predicted diameter distributions of the main tree species of Catalonia. The distributions were calibrated with and without considering the prediction errors of the frequencies of diameter classes. When prediction errors were assumed, the calibration was done with and without making allowance for estimation errors in the stand level calibration variables. Calibrated distributions were more accurate than non-calibrated in terms of sums of different powers of diameters. The set of calibration variables that gave the most accurate results included six stand variables: number of trees per hectare, stand basal area, basal-area-weighted mean diameter, non-weighted mean diameter, median diameter, and basal area median diameter. Of the tested three-variable combinations the best was: number of trees per hectare, stand basal area, and basal-area-weighted mean diameter. Means were more useful calibration variables than medians.

Key words calibration estimation, parameter prediction, Weibull function

Authors' addresses *Palahí*: Centre Tecnològic Forestal de Catalunya, Passeig Lluís Companys, 23, 08010, Barcelona, Spain; *Pukkala*: University of Joensuu, Faculty of Forestry, P.O. Box 111, 80101 Joensuu, Finland; *Trasobares*: Foreco Technologies, Av. Diagonal 416, Estudio 2, Barcelona 08037, Spain

E-mail marc.palahi@ctfc.es

Received 22 November 2005 **Revised** 21 February 2006 **Accepted** 24 May 2006

Available at <http://www.metla.fi/silvafennica/full/sf40/sf403487.pdf>

1 Introduction

Catalonian forests are characterized by heterogeneous stands with a large variation in the spatial distribution of trees, tree species composition, number of stems per hectare, diameter distribution, and vertical structure of the stand. Efficient management planning tools for these forests require growth and yield functions that can produce detailed predictions of stand development under different management schedules. Several researchers have recently developed growth and yield models based on an individual tree approach (e.g. Palahí 2002, Palahí et al. 2003, Palahí and Grau Corbí 2003, Trasobares 2003, Trasobares and Pukkala 2004, Trasobares et al. 2004a, Trasobares et al. 2004b) to address this need in Catalonia. However, when only stand-level inventory data are available, estimating the diameter distribution of trees is required in management planning which uses tree wise growth models (Siitonen 1993). This is the prevailing situation for instance in Finland where ocular inventory is used to assess stand-level characteristics. Only a few actual measurements are taken, most of them being relascope counts of stand basal area. Mean or median diameter is assessed by subjectively selecting a mean or median tree and measuring its diameter (e.g. Mehtätalo 2004).

Various probability density functions, such as normal, gamma, Johnson's SB, Gram-Charlier, beta and Weibull, have been used in describing diameter distributions of forest stands (e.g. Cajanus 1914, Bailey and Dell 1973, Maltamo et al. 1995, Hafley and Schreuder 1977, Maltamo et al. 2000, Palahí et al. 2006b). Palahí et al (2006b) compared the beta, Johnson's SB, Weibull and left-truncated Weibull functions for describing the diameter distributions of basal area and number of stems in forest stands of Catalonia. In this study, the left-truncated Weibull function for stand basal area appeared to be the most accurate function.

In a case where both the stand basal area and the number of stems per hectare are assessed, the predicted distribution is often scaled using one of these variables, and the other is only used as an explanatory variable to predict the parameters of the distribution (Kangas and Maltamo 2003). Using a distribution that provides incorrect values

for known characteristics means wasting information. Therefore, it seems rational to calibrate the predicted distribution so that it is compatible with any additional information on the stand characteristics (Kangas and Maltamo 2000, 2003).

The aim of this study was to examine how the calibration affects the accuracy of the predicted diameter distributions of stands dominated by the main tree species of Catalonia. The method is based on the calibration estimation method of Deville and Särndal (1992), which has already been studied by Kangas and Maltamo (2000, 2003) in Finnish conditions. In the first variant of calibration procedure (referred to as Method 1), prediction errors of class frequencies and estimation errors of stand level calibration variables were not taken into account. Method 2 considered the prediction errors but not estimation errors in field-assessed calibration variables, and Method 3 took into account both the prediction and estimation errors.

Since there were no models available for predicting the diameter distribution function that has been found to be the best in Catalonia (Palahí et al. 2006b), new parameter prediction models were developed for the main tree species of Catalonia based on the left-truncated Weibull function for basal area (Palahí et al. 2006b).

2 Materials and Methods

2.1 Material

The data were provided by the Spanish National Forest Inventory (ICONA, 1993a, ICONA, 1993b, ICONA, 1993c and ICONA, 1993d). They consisted of a systematic sample of permanent plots distributed on a square grid of 1 km, with a 10-year re-measurement interval. From the inventory plots over the whole of Catalonia, all plots of the first inventory (1989–1990) with at least 20 trees were used in this study (see Table 1). This resulted in 3284 plots, the main characteristics of which can be seen in Table 1. One of the eight most common tree species of Catalonia was dominant in 2787 plots. The plots represented all the different stand structures that can be found in Catalonian forests. Most stands were naturally regenerated.

Table 1. Mean, standard deviation (S.D) and range of some characteristics of the study plots. *N* is number of trees (ha⁻¹), *G* is stand basal area (m²ha⁻¹), *D_q* is quadratic mean diameter (cm), *Ele* is elevation (m a.s.l.), and *b* and *c* are Weibull parameters.

Species and no of plots		<i>N</i>	<i>G</i>	<i>D_q</i>	<i>Ele</i>	<i>b</i>	<i>c</i>
<i>Pinus sylvestris</i> 811	Mean	1001	25.2	19.1	1145	24.45	3.68
	S.D	488	8.5	5.0	353	7.56	1.39
	Min	170	12.9	9.3	200	1.12	0.34
	Max	3438	56.8	46.0	1900	61.25	10.57
<i>Pinus uncinata</i> 299	Mean	893	29.1	21.8	1836	28.81	3.46
	S.D	480	9.6	5.4	184	8.08	1.23
	Min	207	13.8	12.4	1200	12.95	1.15
	Max	2865	61.4	40.9	2300	60.89	7.52
<i>Pinus pinea</i> 110	Mean	814	25.8	20.8	180	27.23	4.11
	S.D	332	8.3	4.1	148	5.08	1.71
	Min	325	12.6	13.5	0	13.62	0.71
	Max	2242	57.1	32.8	700	38.59	8.55
<i>Pinus halepensis</i> 320	Mean	1036	20.7	16.7	348	21.05	3.68
	S.D	400	5.3	3.6	216	5.23	1.25
	Min	325	13.1	9.9	0	5.47	0.98
	Max	2642	53.3	33.6	1000	39.64	9.68
<i>Pinus nigra</i> 459	Mean	1232	15.9	15.9	674	19.91	3.60
	S.D	560	3.7	3.7	206	5.86	1.23
	Min	220	10.0	10.0	0	1.00	0.36
	Max	4230	36.9	36.9	1500	48.00	7.58
<i>Abies alba</i> 80	Mean	753	37.3	27.0	1553	40.47	2.94
	S.D	408	12.9	8.0	216	13.55	1.05
	Min	186	16.8	13.6	700	8.14	0.70
	Max	2715	72.4	55.4	2000	80.00	5.44
<i>Quercus ilex</i> 482	Mean	1593	21.2	13.5	545	15.18	2.47
	S.D	585	5.3	2.5	277	5.69	1.10
	Min	420	12.2	9.4	0	1.00	0.46
	Max	3724	41.0	29.4	1300	38.07	6.48
<i>Quercus suber</i> 226	Mean	1046	22.4	17.5	257	23.18	2.83
	S.D	457	6.1	4.1	148	7.05	1.06
	Min	271	12.9	11.1	0	4.70	0.59
	Max	2564	45.2	35.8	800	47.25	6.64
All species 3284	Mean	1119	24.1	17.8	858	22.85	3.34
	S.D	552	8.2	5.3	548	8.98	1.35
	Min	170	12.0	9.3	0	1.00	0.34
	Max	4230	72.4	55.4	2300	80.00	10.57

The sampling method of the national forest inventory uses concentric circular plots in which the plot radius depends on the tree’s diameter at the breast height (dbh, 1.3 m): 5 m radius is used for trees with dbh between 7.5 and 12.49 cm; 10 m for 12.5–22.49 cm; 15 m for 22.5–42.49 cm; and 25 m for dbh greater than or equal to 42.5 cm. Because of the use of variable-radius concentric plots, the number of stems and the basal area per hectare represented by the tree were calculated for every measured tree. The fitting of the diameter distributions used these frequencies rather than assuming an equal frequency of every measured tree.

2.2 Fitting the Diameter Distributions

The probability density function of the left-truncated Weibull is (Zutter et al. 1986)

$$f(d) = \left(\frac{c}{b}\right) \left(\frac{d}{b}\right)^{c-1} \exp\left[\left(\frac{t}{b}\right)^c - \left(\frac{d}{b}\right)^c\right] \quad (1)$$

where *t* is the truncation diameter, *d* is dbh, and *b* and *c* are parameters. The truncation diameter *t* was taken as 7.5 cm, which is the smallest diameter measured on the plots used as study material. Parameters *b* and *c* were estimated by maximizing the following log-likelihood function:

$$\ln L = (c-1) \sum_{i=1}^n g_i \ln(d_i) - G(c-1) \ln(b) + G \ln\left(\frac{c}{b}\right) - \sum_{i=1}^n g_i \left(\frac{d_i}{b}\right)^c + G \left(\frac{t}{b}\right)^c \quad (2)$$

where L is the likelihood function, n is the number of trees on the plot, g_i is the basal area represented by tree i (m^2ha^{-1}) and G is the total basal area of trees (m^2ha^{-1}). The IMSL library routine DBCONF (IMSL... 1997) was employed to maximize the logarithm of the likelihood function.

2.3 Modelling the Parameters of Distributions

The estimated parameters were regressed using mean stand characteristics as explanatory variables. Species-specific models were fitted for the eight most common species (Table 1). In addition, a general model for all species was fitted, and it was used when none of the eight major species was the dominant species (accounted for at least 50% of stand basal area) of the plot. Due to the fact that the modelling data only seldom included stand age, and since the goal was to develop models that could be implemented in any-aged stands, variables such as site index or stand age were not used as explanatory variables. Instead, after graphical analyses, stand basal area (G), number of trees per hectare (N) and their transformations and combinations (like the quadratic mean diameter, $Dq = \sqrt{(40000/\pi G / N)}$) together with geo-topographical variables available for the plots (elevation, slope, aspect, etc) were used as potential predictors. Out of the variables of the latter group, only elevation was a significant predictor. Linear regression analysis and the ordinary least squares (OLS) technique were used to fit the models. All predictors had to be significant at the 0.05 level and the residuals had to indicate an unbiased model.

2.4 Calibrating the Diameter Distribution

The effect of calibration was tested with the predicted diameter distributions. Parameters b and c of the truncated Weibull for the diameter distribution of stand basal area were predicted from N , G and elevation using models developed

in this study. The same 3284 plots that were used to fit the parameter prediction models were used to test the calibration alternatives. It was assumed that the empirical distributions of the plots represent true diameter distributions of Catalanian stands. It was supposed that although the individual distributions contain sampling error, all those distributions are possible in Catalanian forests as true population values. This is justified taking into account the high variability of stand structure due to partial fire damages, irregular cutting systems, multiple tree species and variable growing conditions.

Once the parameters of the distribution function had been predicted, the function was used to calculate the frequencies of trees in different diameter classes. The lower limit of the first diameter class was equal to the truncation diameter (7.5 cm). As the Weibull distribution has no upper limit, it was assumed that the maximum diameter class is the one beyond which the predicted frequency of the class falls below 0.1 trees per hectare. The frequencies obtained from the truncated Weibull function represented the basal areas of diameter classes. They were converted into number of trees by dividing the frequency by the basal area of the class mid-point tree.

The class frequencies obtained in this way were calibrated using the following variable combinations:

- 1) Number of trees per hectare (N)
- 2) Stand basal area (G)
- 3) N and G
- 4) N , G , and basal-area-weighted mean diameter (D_g)
- 5) N , G , and arithmetical mean diameter (D)
- 6) N , G , and basal-area-median diameter (D_{gM})
- 7) N , G , and (frequency) median diameter (D_M)
- 8) N , G , D_g , D , D_{gM} , and D_M

In the first two cases, the class frequencies were only scaled so that the total number of trees or the total basal area of the classes equalled with N or G . In the third case, when Method 1 was used, the class frequencies were calibrated by solving the

following optimization problem with the simplex method (see Deville and Särndal 1992, Kangas and Maltamo 2000):

$$\min \sum_{i=1}^I (s_i^- + s_i^+) \tag{3}$$

subject to

$$w_i + s_i^- - s_i^+ = f_i, \quad i = 1, \dots, I \tag{4}$$

$$\sum_{i=1}^I g_i w_i = G \tag{5}$$

$$\sum_{i=1}^I w_i = N \tag{6}$$

where s_i^+ and s_i^- measure how much the calibrated frequency of diameter class i (w_i) exceeds (s_i^+) or falls short of (s_i^-) the non-calibrated frequency (f_i); I is the number of diameter classes; and g_i, w_i, f_i and d_i are, respectively, the tree basal area (m^2), calibrated frequency, non-calibrated frequency and mid-point diameter (cm) of diameter class i . One-cm diameter classes were used with Method 1.

The calibration adjusted the frequencies of diameter classes so that the total number of trees per hectare (N), stand basal area (G), when calculated from the mid-point trees of diameter classes, agreed with the measured values of these characteristics. When the mean diameters were used as additional calibration variables the following constraints were added to the problem formulation (Kangas and Maltamo 2000, Pukkala and Miina 2005):

$$D_g : \sum_{i=1}^I w_i d_i g_i = G D_g \tag{7}$$

$$D : \sum_{i=1}^I w_i d_i = N D \tag{8}$$

$$D_{gM} : \sum_{i=1}^I \gamma_i w_i g_i = G / 2 \tag{9}$$

$$\text{where } \gamma_i = \begin{cases} 1, d_i < D_{gM} \\ 1/2, d_i = D_{gM} \\ 0, d_i > D_{gM} \end{cases}$$

$$D_M : \sum_{i=1}^I \gamma_i w_i = N / 2 \tag{10}$$

$$\text{where } \gamma_i = \begin{cases} 1, d_i < D_M \\ 1/2, d_i = D_M \\ 0, d_i > D_M \end{cases}$$

In the last problem, where N, G, D_g, D, D_{gM} , and D_M were all used to calibrate the distribution, all the constraints were simultaneously included in the problem formulation.

When Method 2 was used in calibration, the prediction errors of class frequencies were taken into account as proposed by Mehtätalo (2004). The objective function was now:

$$\min \sum_{i=1}^I \frac{1}{\sigma_i} (s_i^- + s_i^+) \tag{11}$$

where σ_i is the standard deviation of the prediction error of tree frequency in diameter class i . The prediction errors were calculated for 5-cm diameter classes using the whole study material of 3284 plots (Fig. 1A). Therefore, Method 2 used 5-cm diameter classes also in calibration. The prediction errors were used to calculate the relative RMSEs (root mean square errors) of the class frequencies (Fig. 1B). These were used to derive the standard deviations (σ_i) in Eq. 11 (predicted frequency \times relative RMSE). The constraints in Method 2 were the same as in Method 1.

Method 3 was otherwise similar to Method 2 except that the assumed estimation errors in calibration variables were also taken into account. In the absence of information from Spain, Finnish studies were consulted (Table 2 in Haara 2002) and the following relative standard errors were assumed: 30% for N , 20% for G and 15% for all means and medians. Method 3 is the way in which calibration estimation should be used in practical calculations which are based on ocular stand inventory and predicted diameter distribution. It was therefore made more realistic also in terms of prediction errors of class frequencies; they were halved from the values in Fig. 1B because the values in Fig. 1B are in reality only partly due to prediction errors while the other part is sampling error. Method 3 was used only for cases in which there were more than one calibration variable. As preliminary tests suggested that medians are less useful calibration variables than means, medians

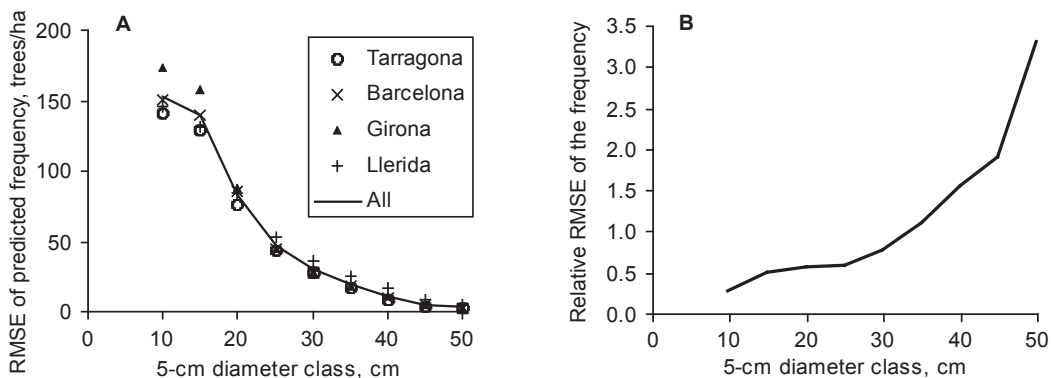


Fig. 1. Root of mean squared error (RSME) of the predicted frequencies of 5-cm diameter classes in the four provinces of Catalonia (A) and the relative RMSE (RMSE/mean prediction) in the whole study material (B).

were not used with Method 3.

When N and G are used as calibration variables with Method 3, the objective function and the constraints for N and G become as follows (the other constraints are the same as before):

$$\min \sum_{i=1}^I \frac{1}{\sigma_i} (s_i^- + s_i^+) + \frac{1}{\sigma_N} (N^- + N^+) + \frac{1}{\sigma_G} (G^- + G^+) \quad (12)$$

$$\sum_{i=1}^I g_i w_i + G^- - G^+ = G \quad (13)$$

$$\sum_{i=1}^I w_i + N^- - N^+ = N \quad (14)$$

where N^+ , N^- , G^+ and G^- measure how much the number of trees or stand basal area exceeds (N^+ , G^+) or falls short of (N^- , G^-) the field assessed value (N , G). The other calibration problems were formulated in the corresponding way.

2.5 Testing the Alternative Calibration Methods

Some of the goal programming problems that were formulated to calibrate the diameter distributions were infeasible. When Method 1 was used, the last problem that included 6 calibration variables could not be solved in about 0.5% of

the plots and the problem with N , G and D_g was infeasible in about 10 plots. The plots that had at least one infeasible calibration problem were not used in the comparisons of calibration variables. With Method 2, which used 5-cm diameter classes, the number of adjustable frequencies was small. Therefore, in about 20% of the plots at least one 3-variable problem was infeasible. As many as 85% of the problems were not feasible when there were 6 calibration variables. The plots in which there were no infeasible 3-variable problems were used to compare combinations of 1, 2 or 3 calibration variables, and with 6 calibration variables those plots were used which could be calibrated for all 6 variables. All problems were feasible with Method 3 with a consequence that the comparisons were based on the whole material on 3284 plots.

The calibration methods were evaluated in the same way as in Maltamo (1997), by calculating the relative biases and RMSEs (square root of mean squared error) for various diameter sums:

$$\%RMSE^c = 100 \times \sqrt{\frac{\sum_{j=1}^N (D_j^c - \hat{D}_j^c)^2}{N-1}} \times \frac{1}{\sum_{j=1}^N \hat{D}_j^c / N} \quad (15)$$

$$\%Bias^c = 100 \times \frac{\sum_{j=1}^N (D_j^c - \hat{D}_j^c)}{N} \times \frac{1}{\sum_{j=1}^N \hat{D}_j^c / N} \quad (16)$$

Table 2. Coefficient of determination (R^2), and absolute and relative RMSE for prediction models of parameter b and c of the truncated Weibull distribution.

Species	Parameter	R^2	RMSE	RMSE%
<i>Pinus sylvestris</i>	b	0.857	8.140	33.3
	c	0.245	1.475	40.1
<i>Pinus uncinata</i>	b	0.846	10.010	34.7
	c	0.200	1.213	35.1
<i>Pinus pinea</i>	b	0.631	9.460	34.7
	c	0.523	1.388	33.8
<i>Pinus halepensis</i>	b	0.889	3.010	14.3
	c	0.155	1.315	35.7
<i>Pinus nigra</i>	b	0.872	4.380	22.0
	c	0.130	1.324	36.8
<i>Abies alba</i>	b	0.842	28.770	71.1
	c	0.252	0.816	27.8
<i>Quercus ilex</i>	b	0.810	6.140	40.4
	c	0.053	1.148	46.5
<i>Quercus suber</i>	b	0.876	6.150	26.5
	c	0.315	0.768	27.1
All species	b	0.869	11.484	50.3
	c	0.215	1.442	43.1

where c is the power to which the diameter is raised, D_j^c is the empirical diameter sum of plot j with power c , \hat{D}_j^c is the corresponding diameter sum calculated from the predicted distribution, and N is the number of plots. The empirical diameter sums for plot j were calculated from

$$D_j^c = \sum_{i=1}^{n_j} n_{ij} d_{ij}^c \tag{17}$$

where n_j is the number of trees measured on plot j , n_{ij} is the number of trees per hectare represented by tree i , and d_{ij} the diameter of tree i of plot j . Power c varied from zero to four. Power zero yields the total number of trees per hectare. The second, third and fourth powers of diameter approximate the stand basal area, stand volume and the economic value of a stand, respectively (Maltamo et al. 1995). To avoid discrepancies caused by other sources of error (e.g. height and volume models) diameter sums were used as the comparison criteria instead of tree volumes (Maltamo et al. 1995).

3 Results

3.1 Parameter Prediction Models

The coefficients of the models for predicting the parameters of the two-parameter truncated Weibull function for the main forest tree species in Catalonia were significant (Tables 3 and 4). The t -values of all parameter estimates were greater than two. N , G , D_q , and transformations of these variables were the main predictors in the parameter prediction models. Elevation (Ele) or squared elevation (Ele^2) was a significant predictor in the common model, and in the models for *Pinus sylvestris*, *P. uncinata* and *Quercus ilex* (Tables 3 and 4).

The model efficiency (R^2), bias and the absolute and relative RMSE for the prediction models of parameters b and c are shown in Table 2. The model efficiency (R^2) and precision (RMSE%) were much lower for the models of parameter c than for the models of parameter b . No serious bias was found in any of the parameter prediction models. The precision (RMSE%) of the model for b ranged from 14% in *Pinus halepensis* to 71% in *Abies alba*. The precision of the model for c was the highest in *Quercus suber* (27%) and lowest in *Q. ilex* (46%).

3.2 Effect of Calibration on the Predicted Distributions

Clear differences in the performance of calibration alternatives were found (Tables 5, 6 and 7). The total ranks for bias and precision summarize the overall accuracy of the tested combinations of calibration variables in terms of the diameter sums used as comparison criteria. Only small differences were found in the ranking of sets of calibration variables between Methods 1 and 2 (Tables 5 and 6) and Method 3 also resulted in nearly similar ranking of the tested combinations of calibration variables (Table 7).

The least accurate results were obtained when using only N or G for scaling the predicted diameter distributions. The combination that gave the most accurate results was the one that included most stand variables: N , G , D_g , D , D_{gM} , and D_M

Table 3. Regression models for the scale (b) parameter of the truncated Weibull distribution. G is stand basal area (m^2ha^{-1}), N is number of trees ha^{-1} , D_q is quadratic mean diameter (cm), and Ele is elevation (m a.s.l.) of the plot.

Species	G	N	D_q	$Ln(N)$	$Ln(G)$	D_q^2	D_q^3	Ele	Ele^2	Const.
<i>Pinus sylvestris</i>	-	-	-	-16.46	16.80	-2.87×10^{-2}	6.32×10^{-4}	-6.40×10^{-3}	3.52×10^{-6}	90.98
<i>Pinus uncinata</i>	-	-	-	-11.11	11.72	-	1.78×10^{-4}	-	1.17×10^{-6}	57.78
<i>Pinus pinea</i>	-0.50	-	-	-10.07	24.10	-	-	-	-	29.84
<i>Pinus halepensis</i>	-	-3.48×10^{-3}	-	-8.45	13.44	-	-	-	-	42.34
<i>Pinus nigra</i>	-	-	-	-10.71	11.35	7.26×10^{-3}	-	-	-	58.44
<i>Abies alba</i>	-0.78	-	1.479	-	32.26	-	-	-	-	-85.01
<i>Quercus ilex</i>	-	-	-	-	0.17	0.17	-4.27×10^{-3}	5.15×10^{-3}	-3.88×10^{-6}	-6.58
<i>Quercus suber</i>	-	-	-	-19.92	20.06	-1.45×10^{-2}	-	-	-	102.73
All species	-	-4.17×10^{-3}	-	-5.71	12.36	1.29×10^{-2}	-	1.63×10^{-3}	-7.57×10^{-7}	23.39

Table 4. Regression models for the shape (c) parameter of the truncated Weibull distribution. G is stand basal area (m^2ha^{-1}), N is number of trees ha^{-1} , D_q is quadratic mean diameter (cm), and Ele is elevation (m a.s.l.) of the plot.

Species	G	N	D_q	$Ln(N)$	$Ln(G)$	D_q^2	D_q^3	Ele	Ele^2	Const.
<i>Pinus sylvestris</i>	5.48×10^{-2}	-	-	-2.07	-	-	-3.61×10^{-5}	2.26×10^{-3}	-1.35×10^{-6}	16.08
<i>Pinus uncinata</i>	-	1.44×10^{-3}	0.43	-	-1.70	-	-1.32×10^{-4}	-	-4.17×10^{-7}	1.44
<i>Pinus pinea</i>	0.34	-	0.26	-	-9.72	-	-	-	-	20.82
<i>Pinus halepensis</i>	-	-	-	-1.26	-	-	-	-	-	12.36
<i>Pinus nigra</i>	-	1.04×10^{-3}	-	-2.20	-	-	-	-	-	17.77
<i>Abies alba</i>	7.15×10^{-2}	-	-0.16	-3.20	-	-	-	-	-	25.62
<i>Quercus ilex</i>	-	9.62×10^{-4}	-	-1.97	-	-	-	-	-	15.37
<i>Quercus suber</i>	-	-	-	-0.69	-	-	6.34×10^{-5}	-	-	7.16
All species	5.74×10^{-2}	1.05×10^{-3}	-	-3.39	-	-2.57×10^{-3}	-	1.07×10^{-3}	-6.63×10^{-7}	24.87

Table 5. Relative biases and RMSEs of different diameter sums that measure the difference between empirical and calibrated predicted distributions when the prediction errors of class frequencies are not considered (Method 1). D_g is the basal-area-weighted mean diameter (cm), D is mean diameter (cm), D_{gM} is the basal area median diameter (cm), D_M is the median diameter (cm), n_j is the frequency (number of trees per hectare) and d_j the mid-point diameter of class i , and R stands for rank.

Calibration variables	Σn_j	R	$\Sigma n_j d_j$	R	$\Sigma n_j d_j^2$	R	$\Sigma n_j d_j^3$	R	$\Sigma n_j d_j^4$	R	Rank sum	Total rank
Bias, %												
N	0.000	1	0.592	7	1.748	8	2.167	8	-0.433	2	26	7
G	2.631	8	1.109	8	0.000	1	-1.659	7	-5.797	8	32	8
$N+G$	0.000	1	-0.014	3	0.000	1	-1.079	6	-5.200	7	15	4
$N+G+D_g$	0.000	1	0.271	5	0.000	1	0.000	1	1.525	3	11	2
$N+G+D$	0.000	1	0.000	1	0.000	1	1.071	5	4.042	5	13	3
$N+G+D_{gM}$	0.000	1	-0.318	6	0.000	1	-0.348	3	-3.634	4	15	4
$N+G+D_M$	0.000	1	-0.127	4	0.000	1	-0.705	4	-4.118	6	16	6
$N+G+D_g+D +D_{gM}+D_M$	0.000	1	0.000	1	0.000	1	0.000	1	0.248	1	5	1
RMSE, %												
N	0.000	1	6.724	7	12.588	8	25.932	8	65.780	8	32	8
G	17.784	8	8.899	8	0.000	1	15.416	7	52.123	7	31	7
$N+G$	0.000	1	3.540	6	0.000	1	13.598	6	47.552	6	20	6
$N+G+D_g$	0.000	1	0.964	3	0.000	1	0.000	1	7.947	2	8	2
$N+G+D$	0.000	1	0.000	1	0.000	1	5.114	3	26.298	3	9	3
$N+G+D_{gM}$	0.000	1	2.978	5	0.000	1	11.606	5	40.436	5	17	5
$N+G+D_M$	0.000	1	2.552	4	0.000	1	10.584	4	38.484	4	14	4
$N+G+D_g+D +D_{gM}+D_M$	0.000	1	0.000	1	0.000	1	0.000	1	3.104	1	5	1

Table 6. Relative biases and RMSEs of different diameter sums that measure the difference between empirical and calibrated distributions when the prediction errors of class frequencies are taken into account (Method 2). D_g is the basal-area-weighted mean diameter (cm), D is mean diameter (cm), D_{gM} is the basal area median diameter (cm), D_M is the median diameter (cm), n_j is the frequency (number of trees per hectare) and d_j the mid-point diameter of class i , and R stands for rank.

Calibration variables	Σn_j	R	$\Sigma n_j d_j$	R	$\Sigma n_j d_j^2$	R	$\Sigma n_j d_j^3$	R	$\Sigma n_j d_j^4$	R	Rank sum	Total rank
Bias, %												
N	0.000	1	1.810	8	2.745	8	-2.173	6	-1.776	3	26	8
G	-0.756	8	0.143	4	0.000	1	-1.617	5	-6.289	6	24	6
$N+G$	0.000	1	0.945	7	0.000	1	-3.256	8	-10.037	8	25	7
$N+G+D_g$	0.000	1	-0.082	3	0.000	1	0.000	1	-0.701	2	8	2
$N+G+D$	0.000	1	0.000	1	0.000	1	-0.611	3	-3.725	4	10	3
$N+G+D_{gM}$	0.000	1	0.162	5	0.000	1	-1.007	4	-4.786	5	16	4
$N+G+D_M$	0.000	1	0.741	6	0.000	1	-2.673	7	-8.447	7	22	5
$N+G+D_g+D+D_{gM}+D_M$ ¹⁾	0.000	1	0.000	1	0.000	1	0.000	1	-0.266	1	5	1
RMSE, %												
N	0.000	1	4.825	7	9.208	8	20.161	8	56.743	8	32	8
G	9.826	8	6.032	8	0.000	1	13.328	7	49.785	6	30	7
$N+G$	0.000	1	3.047	6	0.000	1	13.262	6	51.342	7	21	6
$N+G+D_g$	0.000	1	0.914	3	0.000	1	0.000	1	10.951	2	8	2
$N+G+D$	0.000	1	0.000	1	0.000	1	4.436	3	26.760	3	9	3
$N+G+D_{gM}$	0.000	1	3.032	5	0.000	1	11.592	5	42.361	5	17	5
$N+G+D_M$	0.000	1	2.473	4	0.000	1	9.962	4	40.397	4	14	4
$N+G+D_g+D+D_{gM}+D_M$ ¹⁾	0.000	1	0.000	1	0.000	1	0.000	1	2.368	1	5	1

¹⁾ Based on 567 plots that could be calibrated for all six stand variables. The other results are based on 2899 plots.

Table 7. Relative biases and RMSEs of different diameter sums that measure the difference between empirical and calibrated distributions when the prediction errors of class frequencies and the estimation errors of calibration variables are taken into account (Method 3). D_g is the basal-area-weighted mean diameter (cm), D is mean diameter (cm), n_j is the frequency (number of trees per hectare) and d_j the mid-point diameter of class i , and R stands for rank.

Calibration variables	Σn_j	R	$\Sigma n_j d_j$	R	$\Sigma n_j d_j^2$	R	$\Sigma n_j d_j^3$	R	$\Sigma n_j d_j^4$	R	Rank sum	Total rank
Bias, %												
$N+G$	-0.021	1	0.155	3	0.000	1	-2.035	4	-8.910	4	13	4
$N+G+D_g$	0.158	4	0.396	4	0.133	4	-0.012	1	0.336	2	15	3
$N+G+D$	-0.116	3	0.000	1	0.019	2	0.660	3	1.610	3	12	2
$N+G+D_g+D$	-0.057	2	0.013	2	0.101	3	0.160	2	-0.283	1	10	1
RMSE, %												
$N+G$	2.783	3	3.859	4	0.000	1	17.149	4	77.005	7	19	4
$N+G+D_g$	4.486	4	3.734	3	2.198	4	0.640	1	16.609	2	14	3
$N+G+D$	1.181	1	0.000	1	1.124	2	6.931	3	40.413	3	10	1
$N+G+D_g+D$	2.388	2	1.340	2	1.779	3	6.094	2	8.772	1	10	1

with Methods 1 and 2, and N , G , D_g and D with Method 3. This was followed in the ranking by combination N , G , and D_g . The third position in the overall accuracy was for N , G , and D . The non-weighted mean diameter (D) was almost equally useful calibration variable as the basal-area-weighted mean (D_g) in terms of rank sum (Tables 5 and 6), sometimes even better (Table 7). However, when sums of diameters raised to the third or fourth power was looked at, D_g was clearly superior to D . Using a median diameter (D_M or D_{gM}) as a calibration variable did not improve much the accuracy of estimating or predicting the diameter distributions when compared to the situation where only G and N were used.

Accurate results for both the stand basal area and the total number of trees required that both N and G were used in calibration. Using non-weighted mean diameter (D) as a third calibration variable removed all errors from the sums of first powers of diameter, also with Method 3 in which exact agreement was not required. The use of D_g with N and G enabled accurate results for the sum of the third powers of diameter. An accurate estimation of the sum of fourth powers required the use of all six or four calibration variables, but combination N , G , and D_g was also quite good.

Examples of the ability of different sets calibration variables to improve the predicted diameter distributions of six plots with Method 1 are shown in Fig. 2. There were clear differences between the

best set (six stand variables) and using G as a scaling variable (Fig. 2). Fig. 2 shows that by using six calibration variables (alternative ‘All’ in Fig. 2), uni-modal, bi-modal, descending and irregular distributions could be accurately described.

Fig. 3 shows examples of the effect of calibration method (1–3) on the resulting distribution. Method 3, in which there are no strict constraints for N , G and mean diameters, produces smooth distributions that do not deviate much from the predicted non-calibrated distribution.

4 Discussion

This study developed parameter prediction models for the truncated Weibull function for the diameter distribution of stand basal area of the main forest tree species of Catalonia. It then examined how the calibration affected the accuracy of the predicted distributions. The modelling data used in the study reflected the complexity and heterogeneity of Catalonian forests. The study data encompassed plots of regular and irregular stand structures with unimodal, decreasing, uniform and even multimodal size distributions of trees (see Figs. 2 and 3). In these types of stands, plots with the same basal area and number of stems may have quite different diameter distributions.

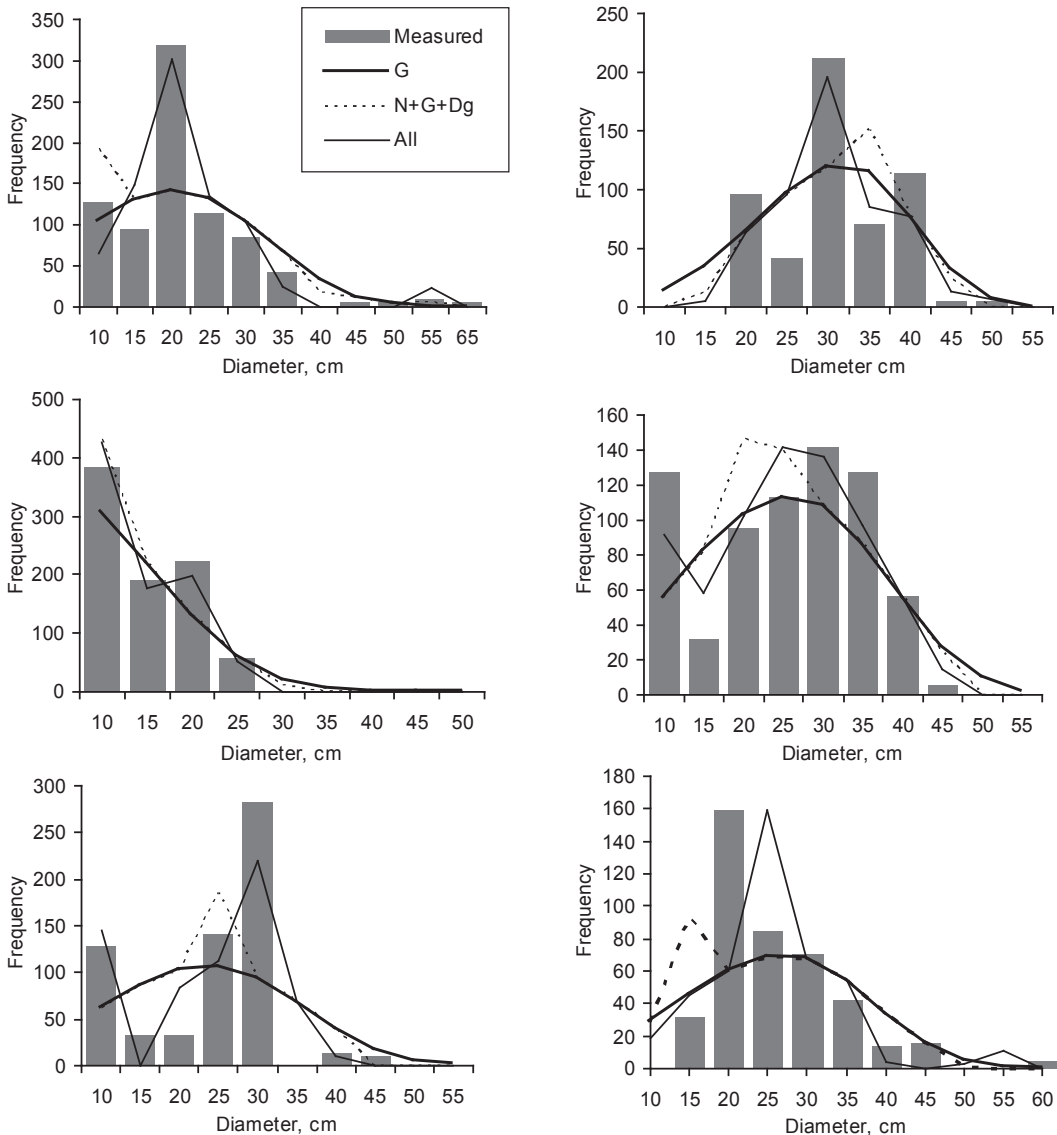


Fig. 2. Examples of measured, predicted (scaled with G), calibrated for N , G and D_g , and calibrated for six variables (All) distributions for six plots of the study material when Method 1 is used in calibration (no prediction error in class frequencies and no estimation error in calibration variables). ‘Frequency’ is the number of trees per hectare of the 5-cm diameter class.

This is reflected in the accuracy of the parameter prediction models. However, since the sampling method was not specifically designed to develop models and estimate diameter distributions, the sample presents some limitations. One limitation was that small trees (<7.5 cm) were not measured individually with a consequence that all our

results concern the functions’ ability to describe the distribution of trees larger than 7.49 cm in diameter. In addition, the plots were rather small and most of them had too few trees to reliably characterize the diameter distribution of the whole stand. Because of this, only plots with at least 20 measured trees were used.

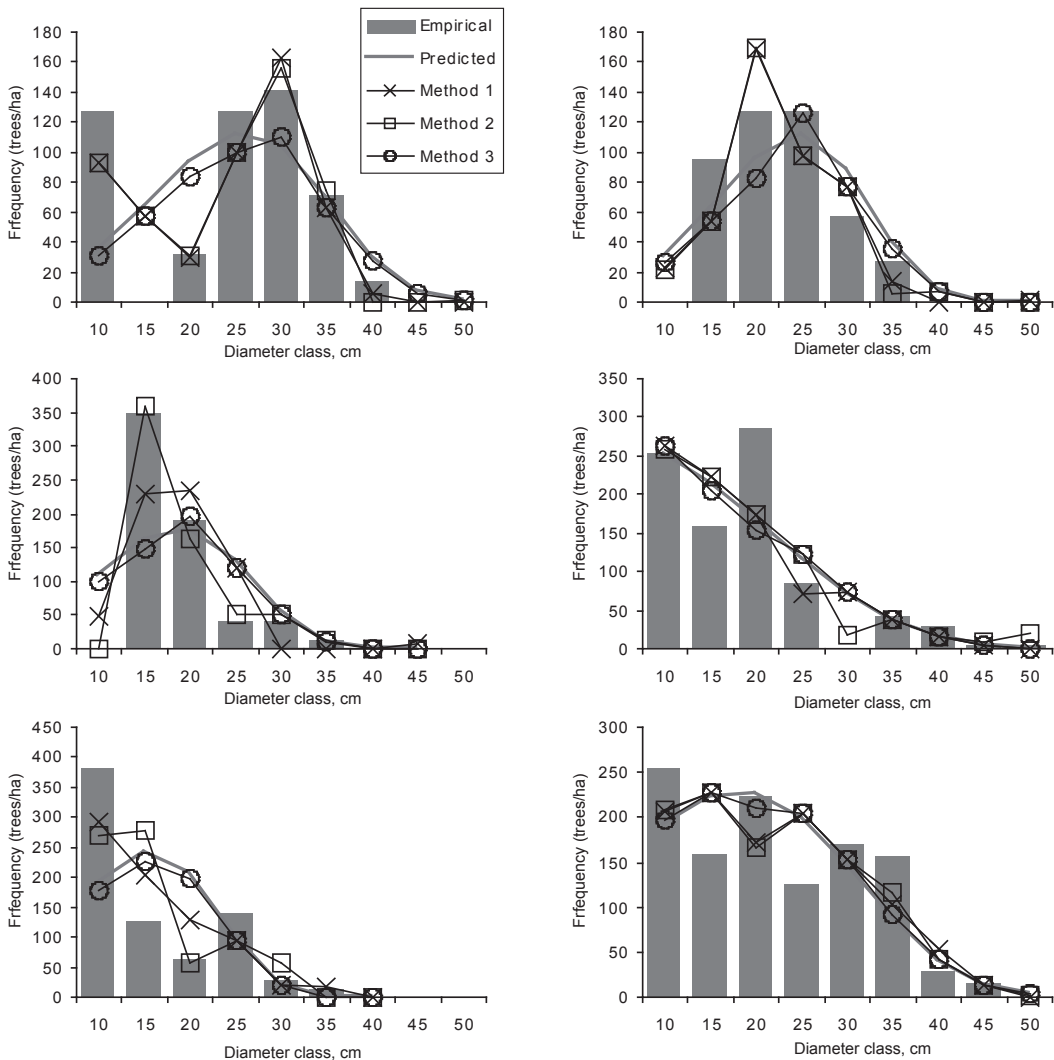


Fig. 3. Examples of measured, predicted (scaled with G), and calibrated with N , G and D_g and D distributions for six plots of the study material when three different methods are used in calibration. Method 1 assumes no prediction errors in class frequencies and no estimation errors in calibration variables. Method 2 assumes prediction errors in class frequencies but no estimation errors in calibration variables. Method 3 assumes prediction errors in class frequencies and estimation errors in calibration variables.

The idea behind the comparison criteria was to study the performance of different calibration alternatives in estimating and predicting variables that correlate with the number of stems, mean diameter, stand basal area, stand volume, and the economic value of the stand, but at the same time avoiding discrepancies caused by other sources of error (e.g. height and volume models). Therefore,

diameter sums were used instead of for instance stand volume of stumpage value of trees.

The sum of the absolute deviations from the non-calibrated frequencies (Eq. 3) was used as the distance measure to be minimised subject to the calibration equations. Kangas and Maltamo (2000) tested several distance functions and obtained the most accurate results for the function

that was used in this study. However, Kangas and Maltamo (2003) concluded that further studies are required to analyse the performance of different distance functions. The problem of the objective function employed in this study is that neighbouring diameter classes can have very unequal calibrated frequencies. The peaks and falls of the true diameter distributions would most probably distribute over more diameter classes than the solutions of our calibration problems suggest. However, the tendency of goal programming to concentrate the changes in tree frequencies on too few diameter classes may have a rather small influence on the calculation results that are based on calibrated distributions. The problem of very unequal frequencies of neighbouring classes is much smaller with calibration Method 3, which allows flexibility in the values of stand variables computed from the calibrated distribution.

Calibrating the diameter distributions proved to be an efficient way of using all available information and generating more accurate diameter distributions in terms of sums of different powers of diameter. These results are in accordance with the study of Kangas and Maltamo (2000) who found that using G , N and D_{gM} to calibrate the distribution produced more accurate results than using only G and N . In our study, the use of medians as additional calibration variables also improved the distribution, but combination G , N and D_{gM} performed clearly worse than G and N with a mean diameter (D or D_g). The order of the means and medians, when one of them was used together with G and N , was (from best to worst): $D_g > D > D_{gM} & D_M$. This suggests that it would be more useful to measure means rather than medians in the field survey. The advantage of medians is that they can be measured more easily than means in circular (D_M) or relascope (D_{gM}) plots.

In our study the RMSE% values for the sum of the third powers of diameter were higher than the RMSE% values in Kangas and Maltamo (2000) for volume when the same calibration equations were used. This might be due to the complex structures (irregular, bimodal, descending) of the stands used as the data of this study. However, at the same time the structural complexity justifies even more the use of calibration to improve the prediction of the diameter distributions of Catalan forest stands.

Calibrating the diameter distribution is an efficient way of using additional information in the calculation of forest inventory results. One advantage of calibration estimation is that the same variables need not to be known from each stand (Kangas and Maltamo 2000). Calibration estimation enables the use of all combinations of stand variables that are collected in stand inventory. The minimum requirement is that those variables that are needed to predict the parameters of the diameter distribution are assessed (G and N for the parameter prediction models of in this study).

In Catalonia, forest inventory is the most expensive task of forest management planning. Usually, many circular, concentric or relascope plots are placed within the compartment, and the diameters of individual trees of the plots are measured and recorded. This study shows that cheaper forest inventory methods could also be used with only G and N measured in all stands, plus mean diameters in heterogeneous stands or in places where more accurate information is required.

Acknowledgments

This study has been conducted within the research project AGL2004-00382 funded by the Ministerio de Educación y Ciencia of Spain, Secretaria de Política Científica y Tecnológica.

References

- Bailey, R.L. & Dell, T.R. 1973. Quantifying diameter distribution with the Weibull-function. *Forest Science* 19(2): 97–104.
- Cajanus, W. 1914. Über die Entwicklung gleichaltriger Waldbestände. Eine statistische Studie. I. *Acta Forestalia Fennica* 3. 142 p.
- Deville, J.C. & Särndal, C.E. 1992. Calibration estimators in survey sampling. *Journal of American Statistical Association* 87: 376–382.
- Haara, A. 2002. Kasvuennusteiden luotettavuuden selvittäminen knn-menetelmällä ja monitavoiteoptimoinnilla. *Metsätieteen aikakauskirja* 3/2002: 391–406. (in Finnish)
- Hafley, W.L. & Schreuder, H.T. 1977. *Statistical distri-*

- butions for fitting diameter and height data in even-aged stands. *Canadian Journal of Forest Research* 4: 481–487.
- ICONA. 1993a. Segundo Inventario Forestal Nacional (1986–1995) Cataluña: Barcelona, MAPA, Madrid.
- ICONA. 1993b. Segundo Inventario Forestal Nacional (1986–1995) Cataluña: Girona, MAPA, Madrid.
- ICONA. 1993c. Segundo Inventario Forestal Nacional (1986–1995) Cataluña: Lleida, MAPA, Madrid.
- ICONA. 1993d. Segundo Inventario Forestal Nacional (1986–1995) Cataluña: Tarragona, MAPA, Madrid.
- IMSL. 1997. Fortran subroutines for mathematical applications. Visual Numerics, Inc. Houston, Texas. 1311 p.
- Kangas, A. & Maltamo, M. 2000. Calibrating predicted diameter distribution with additional information. *Forest Science* 46(3): 390–396.
- & Maltamo, M. 2003. Calibrating predicted diameter distribution with additional information in growth and yield predictions. *Canadian Journal of Forest Research* 33(3): 430–434.
- Maltamo, M. 1997. Comparing basal area diameter distributions estimated by tree species and for the entire growing stock in a mixed stand. *Silva Fennica* 31(1): 53–65
- , Puumalainen, J. & Päivinen, R. 1995. Comparison of beta and Weibull functions for modeling basal area diameter distribution in stands of *Pinus sylvestris* and *Picea abies*. *Scandinavian Journal of Forest Research* 10: 284–295.
- , Kangas, A., Uutera, J., Tornainen, T. & Saranmäki, J. 2000. Comparison of percentile based prediction methods and the Weibull distribution in describing the diameter distribution of heterogeneous Scots pine stands. *Forest Ecology and Management* 133: 263–274.
- Mehtätalo, L. 2004. An algorithm for ensuring the compatibility between estimated percentiles of diameter distribution and measured stand variables. *Forest Science* 50(1): 20–32.
- Palahí, M. 2002. Modelling stand development and optimising the management of even-aged Scots pine forests in north-east Spain. Academic Dissertation. Research notes 143. Faculty of Forestry, University of Joensuu.
- & Grau Corbí, J. 2003. Preliminary site index model and individual-tree growth and mortality models for Black pine (*Pinus nigra* Arn.) in Catalonia (Spain). *Invest. Agr.: Sist. Recur. For.* 12(1): 137–148.
- , Pukkala, T., Miina, J. & Montero, G. 2003. Individual-tree growth and mortality models for Scots pine (*Pinus sylvestris* L.) in north-east Spain. *Annals of Forest Science* 60: 1–10.
- , Trasobares, A. & Pukkala, T. 2006a. Modelling the diameter distribution of the number of trees per hectare for *Pinus sylvestris*, *Pinus nigra* and *Pinus halepensis* forest stands in Catalonia using the Weibull function. Submitted to *Forestry*.
- , Trasobares, A. & Pukkala, T. 2006b. Comparison of Beta, Johnson's SB, Weibull and truncated Weibull functions for modelling the diameter distribution of forest stands in Catalonia (north-east of Spain). Submitted to the *European Journal of Forest Research*.
- Pukkala, T. & Miina, J. 2005. Optimising the management of a heterogeneous stand. *Silva Fennica* 39(4): 525–538.
- Trasobares, A. 2003. Modelling and management of uneven-aged pine forests in Catalonia, north-east Spain. Academic Dissertation. University of Joensuu, Faculty of Forestry, Research Notes 155.
- & Pukkala T., 2004. Using past growth to improve individual-tree diameter growth models for uneven-aged mixtures of *Pinus sylvestris* L. and *Pinus nigra* Arn. in Catalonia, north-east Spain. *Annals of Forest Science* 61(5): 409–417.
- , Pukkala T. & Miina J. 2004a. Growth and yield model for uneven-aged mixtures of *Pinus sylvestris* L. and *Pinus nigra* Arn. in Catalonia, north-east Spain. *Annals of Forest Science* 61(1): 9–24.
- , Tomé M. & Miina J. 2004b. Growth and yield model for *Pinus halepensis* Mill. in Catalonia, north-east Spain. *Forest Ecology and Management* 203(1): 49–62.
- Zutter, B.R., Oderwald, R.G., Murphy, P.A. & Farrar, R.M. 1986. Characterizing diameter distributions with modified data types and forms of the Weibull distribution. *Forest Science* 32(1): 37–48.

Total of 27 references