# The Gompertz Type Stochastic Growth Law and a Tree Diameter Distribution

**PETRAS RUPŠYS, EDMUNDAS PETRAUSKAS*, JUOZAPAS MAŽEIKA AND ROMUALDAS DELTUVAS**
*Institute of Fundamental Sciences & Forest Management Department of Lithuanian University of Agriculture Studentų 11, LT-53361 Akademija, Kaunas district, Lithuania, edmundas.petrauskas@lzuu.lt;*

**Rupšys, P., Petrauskas, E., Mažeika, J. and Deltuvas R.** 2007. The Gompertz Type Stochastic Growth Law and a Tree Diameter Distribution. *Baltic Forestry*, 13 (2): 197–206.

*Abstract*

The study presents a comprehensive stand-level model for the tree diameter distribution. A Gompertz type stochastic logistic growth law is used for describing diameter distribution. Using the Gompertz type stochastic law of tree diameter growth, the age and height dependent probability density function of diameter distribution is obtained. The mean age-diameter, height-diameter growth trends and their variances for the Gompertz type stochastic differential equation are derived. The expected tree diameter distribution is predicted by using the Fokker-Plank equation and stand measurements. The estimates of parameters are performed by the $L^1$ distance procedure. The Weibull, and negative exponential distributions are selected to study their performance to the observations. To evaluate the goodness-of-fit, the absolute discrepancy, Kolmogorov-Smirnov, and Reynolds error index statistics are adapted. In addition, for estimating the goodness-of-fit, the Chi-squared test, pseudo-residuals, and Shapiro-Francia statistic are arranged, and the normal quantile plot is described. To model the diameter distribution, as an illustrative experience, a real data set from repeated measurements on permanent sample plots of pine stands in Dubrava forest district is used. The results are implemented in the symbolic computational language MAPLE.

**Key words:** Weibull, Gompertz, negative exponential, diameter distribution, stochastic growth law, Fokker-Plank equation

## Introduction

Forest as association of geo-bio ingredients has stochastic nature. It consists of nucleus of a biotic communions and abiotic environment. Stand as a community of trees is the main component of the forest. Stand consists of trees with different heights and diameters. Those differences depend on a lot of unsearchable genetic and environmental factors, therefore it leads to consideration that diameter and height of trees are random variables. The theory of probability declares that random variable could be fully described knowing distribution of its values. Distribution function gives us detailed view of random variable while mean value only general view. Distribution functions of diameter, height, and volume are important for forest theory and practice. Those distributions depend on age of stand and site type. Forest type is defined by tree species, age of a stand (age class), site type. Site type is described by mean height of trees.

The main goal of forest inventories is unbiased assessment of the increment and yield of wood. It is evident that stand vise forest inventory due to limited human and financial resources is unrealizable. As a solution sampling methods are used. To get data of known accuracy and confidence, sampling must be organized on a scientific base. Due to reliability and economy of obtained results, it is expedient to approximate sampling data by multiply distribution functions. Tree species diameter distribution functions should describe all inventory unit according to age and site type. Inventory unit for stand wise forest inventory might be forest holdings and all country forests for sample based national inventory.

The study presents a comprehensive stand-level model for the tree diameter distribution that might be used in sample based national inventory.

Diameter is an important element in forest stand, as well as height, basal area, stand volume and number of trees. The distributions of tree diameter in stands describe forest structure and are used for the assessment of stand volume and biomass (Chen 2004), forest biodiversity (Uuttera *et al.* 1995), density management (Newton 1997, Sterba 2004, Newton *et al.* 2005). The idea to describe the distribution of diameter in a stand is the ultimate question for foresters. There are three types of approaches for this purpose (Mehtatalo 2005 and references therein). The first approach is based on a sample of diameters of a stand (Tang *et al. 1*997, Mehtatalo 2005). The second approach is based on parameters prediction in which parameters of a probability density function are predicted from some easily measured stand characteristics (Lindsay *et al*. 1996, Alvarez *et al*. 2002, Newton *et al.* 2005). In the third parameter recovery approach parameters of the probability density function are predicted from some easily measured stand-level variables (Lindsay *et al*. 1996, Cao 2004, Chen 2004, Dieguez-Aranda *et al.* 2005). The last two approaches can be conjoined and called the probability density function method. The probability density function method exploits directly the fact that all observations should come from the same distribution law. There are several ways of determining this theoretical distribution, such as negative exponential distribution (Meyer *et al.* 1943, Leak 1996), Pearson distribution (Schnur 1934), gamma distribution (Nelson 1964), lognormal distribution (Bliss *et al*. 1964, Chen 2004), beta distribution (Clutter *et al*. 1965, Chen 2004), Weibull distribution (Weibull 1951, Bailey *et al.* 1973, Maltamo *et al*. 2000, Zucchini *et al*. 2001, Cao 2004, Chen 2004, Dieguez-Aranda *et al.* 2005, Mehtatalo 2005, Newton *et al.* 2005, Westphal *et al.* 2006, Nord-Larsen *et al.* 2006), Johnson distribution (Tham 1988, Hafley *et al.* 1977, Zucchini *et al.* 2001, Chen 2004), double–normal distribution (Bruchwald 1988), Charlier distribution (Prodan 1953), Pearl-Read distribution (Nelson 1964), de Liocourt distribution (1898), semi logarithmic distribution (Sterba 2004), finite mixture distribution (Zucchini *et al*. 2001, Zasada *et al.* 2005), bivariate distribution (Knoebel *et al*. 1991, Uusitalo *et al*.1998, Tewari *et al*. 1999, Zucchini *et al*. 2001, Li *et al.* 2002), and much more. The history of mathematical modeling of diameter distribution have pointed out that most part of theoretical distributions fits for the pure even-aged stands. Comprehensive tree diameter distribution studies have demonstrated that diameter is not normally distributed (Bliss *et al*. 1964, Mehtatalo 2005, Nord-Larsen *et al.* 2006).

Conventionally, the Weibull law is applied for characterizing the diameter distribution. The probability density function of the Weibull law has the following form (Weibull 1951)

$$p(x)=\begin{cases}\left(\frac{c}{b}\right)\left(\frac{x-a}{b}\right)^{c-1}\exp\left(-\left(\frac{x-a}{b}\right)^{c}\right), x\geq a,\\ 0, x<a,\end{cases} \tag{1}$$

where *a, b, c* are the location, scale, shape parameters, and *x* is tree diameter of breast height. Mostly, as indicated by Cao (2004), the parameters *a, b, c* are estimated by the technique of multiple regression using the predictor stand-level variables such as the stand age in years, the dominant height in meters, the number of trees per ha, and the relative spacing.

The main goal of this study is to deduce the original mixture of the Gompertz type (Gompertz 1825, Ananda *et al.* 1996) transition probability density functions fitted for the modeling of diameter distributions of trees in even-aged, and uneven-aged stands. Our presented method of tree diameter distribution by the age or the height dependent probability density function applies measurements of stand variables such as age, height, diameter at breast height. This age and height dependent tree diameter distribution accumulates additional age, height, species structure information of stands. This method combines the parameter prediction and parameter recovery approaches. The parameters of the distribution function are recovered from the Gompertz type dynamic diameter growth law, which trajectory is governed by a standard Brownian motion (Rupšys 2004, 2005). Because the extent of unknown variability of tree diameter is associated with forest growth processes, we incorporate stochastic structure in tree diameter growth model and hold transitions among diameter classes on the subject of the age and height. Our used Gompertz type stochastic growth law includes biologically relevant nonlinear mechanisms of diameter growth. For the modeling of diameter growth, we used the state space stochastic differential equation approach (Garcia 1983, 1994, 2005, Rupšys 2003, 2005; Matis *et al*. 2003). This approach applies stochastic differential equations and improves characterization of actual diameter growth (Garcia 2005). For the estimation of parameters, the maximum likelihood procedure or $L^1$ distance procedure (Shoi *et al*. 1997, Shoi 1998, Rupšys 2004, 2005) is used. In this paper we apply the $L^1$ distance procedure (Rupšys 2005). However, in practice it is not simple to materialize our presented model of the tree diameter distribution.

For validation, the developed mixture distribution is compared to the Weibull and negative exponential distributions. To evaluate the goodness-of-the-fit between the field data and the estimated distribution, three statistics are calculated: the absolute discrepancy (Westphal 2006), the Kolmogorov-Smirnov (Kubilius 1980), and the Reynolds error index (Reynolds *et al.* 1988). Finally, the Chi-squared test (Nikulin 1973), the pseudo-residuals (Zucchini *et al*. 1999) are examined, and the Shapiro, and Francia statistics *W'* (Shapiro *et al*. 1972), for evaluating the straightness of a normal quantile plot, is calculated.

## Material and methods

### *Data*

The diameter analysis is based on experiments in pine stands of Dubrava forest district. These stands have been measured 5 times on stand variables: age, number of trees per hectare, breast height diameter, trees position co-ordinates, age and height on each tenth tree. The measurements have been conducted in 34 permanent treatment plots, and the initial planting densities are unknown. Plot area is 0.25 ha. The age of stands ranges from 12 to 103 years. The mean of diameter at breast height varies from 2.5 to 51 cm. Approximately 10% of trees in all plots are randomly selected for the height measurement. The observed data of study plots are presented in Figure 1.

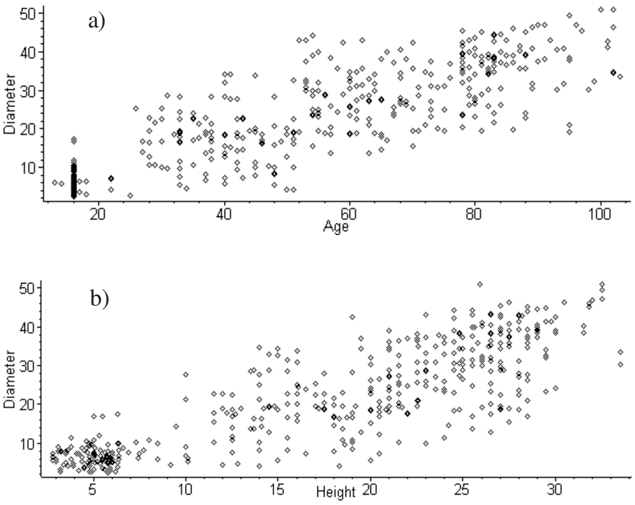



**Figure 1.** Plot of the diameter including data from pine forests at Dubrava district in Lithuania: a) the age dependent, b) the height dependent

***Technique***

Suppose the dynamics of diameter growth is expressed in terms of the Gompertz type stochastic ordinary differential equations (Rupšys 2003)

$$dx(t) = r_1 x(t) \ln \frac{K_1}{x(t)} dt + \sigma_1 x(t) dW(t) \tag{2}$$

$$dx(h) = r_2 x(h) \ln \frac{K_2}{x(h)} dt + \sigma_2 x(h) dW(h) \tag{3}$$

where $r_1$, $r_2$ are the diameter intrinsic growth, $K_1, K_2$ are the diameter carrying capacity and forms a numerical upper bound on the diameter size, and $x(t)$, $x(h)$ are the tree breast height diameter at age $t$ , and height $h$, $\sigma_1$, $\sigma_2$ are the intensity of noises, $W(t)$, $W(h)$ are scalar Brownian motions. This diameter growth model uses two transition functions expressed as the stochastic differential equations, which project the future of tree diameter subject to the age or height. A probability density function of tree diameter distribution can be obtained from the Fokker–Plank, or forward Kolmogorov equation (Gihman *et al.* 1977), which relates the variation of tree diameter with the age (height). The nature of diameter growth allows us to choose the Ito stochastic calculus. Each solution of the stochastic differential equations (2)-(3) describes one path of diameter development. The ensemble of realizations, described by transitional probability density function $p_1(x,t)$ $(p_2(x,h))$, satisfies the corresponding Fokker–Plank equation

$$\frac{\partial p_1(x,t)}{\partial t} = -r_1 \frac{\partial}{\partial x}\left( x \ln \frac{K_1}{x} p_1(x,t) \right) + \frac{\sigma_1^2}{2} \frac{\partial^2}{\partial x^2}\left( x^2 p_1(x,t) \right), \tag{4}$$

$$\frac{\partial p_2(x,h)}{\partial h} = -r_2 \frac{\partial}{\partial x}\left( x \ln \frac{K_2}{x} p_2(x,h) \right) + \frac{\sigma_2^2}{2} \frac{\partial^2}{\partial x^2}\left( x^2 p_2(x,h) \right). \tag{5}$$

The solution of the Fokker–Plank equations (4)-(5) describes the evolution of probability density function of tree diameter through age (height). In the sequel we assume that initial tree diameter is not random and known exactly $x(t_0) = x_0$.

The transformation

$y(t) = \ln(x(t))$ (y(h) = $\ln(x(h))$ )

the diffusion term $\sigma x(t)$ $(\sigma x(h))$ in stochastic model (2)-(3) converts to $\sigma$ and the nonlinear process (2)-(3) transforms into the Ornstein-Uhlenbeck process (Gihman *et al.* 1977, Soboleva *et al.* 2003). So, we can find the age (height) dependent solution of the Fokker-Planck equations (4)-(5). The solution of equation (4)-(5) has the following form

$$p_1(x,t|r_1,K_1,\sigma_1) = \frac{1}{\sigma_1 x \sqrt{\pi\left(1-e^{-2r_1 t}\right)/r_1}} e^{-\frac{r_1\left(\ln x - \ln K_1 + \sigma_1^2/2r_1 - e^{-r_1 t}\ln x_0\right)^2}{\sigma_1^2\left(1-e^{-2r_1 t}\right)}}, \tag{6}$$

$$p_2(x,h|r_2,K_2,\sigma_2) = \frac{1}{\sigma_2 x \sqrt{\pi\left(1-e^{-2r_2 h}\right)/r_2}} e^{-\frac{r_2\left(\ln x - \ln K_2 + \sigma_2^2/2r_2 - e^{-r_2 h}\ln x_0\right)^2}{\sigma_2^2\left(1-e^{-2r_2 h}\right)}}. \tag{7}$$

Using the time (height) dependent probability density functions (6)-(7), we can analyze the behavior of the mean trend $m_t = E(x(t))$ $(m_h = E(x(h)))$ and variance $s_t = V(x(t))$ $(s_h = V(x(h)))$ of tree diameter. When the instantaneous fluctuation is proportional to diameter $x(t)$ $(x(h))$, then the mean trend and variance of tree diameter fulfill the system of ordinary differential equations

$$\begin{cases} \dfrac{dm_t}{dt} = r_1 m_t \ln \dfrac{K_1}{m_t} - \dfrac{r_1 s_t}{2m_t}, \\ \dfrac{ds_t}{dt} = 2r_1 s_t \left( \ln \dfrac{K_1}{m_t} - 2 \right) + \sigma_1{}^2\left(m_t^2 + s_t\right), \end{cases} \tag{8}$$

$$\begin{cases} \dfrac{dm_h}{dh} = r_2 m_h \ln\dfrac{K_2}{m_h} - \dfrac{r_2 s_h}{2m_h}, \\ \dfrac{ds_h}{dh} = 2r_2 s_h\left(\ln\dfrac{K_2}{m_h} - 2\right) + \sigma_2^{\,2}\left(m_h^2 + s_h\right) \end{cases} \qquad (9)$$

The general formal structure of the tree diameter probability density function $p(x,t,h|r,K,\sigma))$ can be defined as a mixture of two probability density functions $p_1(x,t|r_1,K_1,\sigma_1)$, $p_2(x,h|r_2,K_2,\sigma_2)$. Consequently, mixed probability density function $p(x,t,h|r,K,\sigma)$ is expressed as a weighted sum of density functions $p_1(x,t|r_1,K_1,\sigma_1)$, $p_2(x,h|r_2,K_2,\sigma_2)$ in the following form

$$p(x,t,h|r,K,\sigma) = w_1 p_1(x,t|r_1,K_1,\sigma_1) + w_2 p_2(x,h|r_2,K_2,\sigma_2) \qquad (10)$$

where $0 \le w_1, w_2 \le 1$, $w_1 + w_2 = 1$

Now we consider the estimate of parameters of probability density function $p(x,t;h|r,K,\sigma)$ using the Gompertz stochastic diameter growth models (2)-(3). The parameters of the stochastic differential equations (2)-(3) can be estimated by the maximum likelihood procedure or the $L^1$ distance procedure (Rupšys 2004, 2005).

The maximum likelihood function of the observed data $\{(t_i,x_i),\ i=1,2,...,n\}$ ( $\{(h_i,x_i),\ i=1,2,...,n\}$ ), has the following form (Rupšys 2004)

$$L_1(r_1,K_1,\sigma_1) = -\frac{1}{2}\sum_{i=2}^{n'}\left[\frac{(\ln x(t_i) - E_{i-1})^2}{V_{i-1}} + \ln(2\pi V_{i-1})\right] + \sum_{i=0}^{n}\ln\left(\frac{d\phi(x_i)}{dx}p_i(x_i)\right) \qquad (11)$$

$$L_2(r_2,K_2,\sigma_2) = -\frac{1}{2}\sum_{i=2}^{n'}\left[\frac{(\ln x(h_i) - E_{i-1})^2}{V_{i-1}} + \ln(2\pi V_{i-1})\right] + \sum_{i=0}^{n}\ln\left(\frac{d\phi(x_i)}{dx}p_i(x_i)\right) \qquad (12)$$

where $n'$ is the number of the different ages (heights) of the observed data, $x(t_i) = x_i'$ ( $x(h_i) = x_i'$ ), $x_i'$ is the average of the observed data at time $t_i$, $(h_i)\Delta t_{i-1} = t_i - t_{i-1}$, $t_0 < t_1 < ... < t_{n'}$ ($\Delta h_{i-1} = h_i - h_{i-1}$, $h_0 < h_1 < ... < h_{n'}$ ), $p_i(.)$ is the density function of the *i-th* observation (in the sequel we assume that it follows the normal distribution),

$$E_{i-1} = \ln x(t_{i-1}) + \left(\ln x(t_{i-1}) - \ln K_1 + \frac{1}{2r_1}\sigma_1^2\right)\left(e^{-r_1\Delta t_{i-1}} - 1\right)$$

$$E_{i-1} = \ln x(h_{i-1}) + \left(\ln x(h_{i-1}) - \ln K_2 + \frac{1}{2r_2}\sigma_2^2\right)\left(e^{-r_2\Delta h_{i-1}} - 1\right)$$

$$V_{i-1} = \frac{\sigma_1^2}{2r_1}\left(1 - e^{-2r_1\Delta t_{i-1}}\right)$$

$$V_{i-1} = \frac{\sigma_2^2}{2r_2}\left(1 - e^{-2r_2\Delta h_{i-1}}\right)$$

The maximum likelihood procedure may be fail when we own a few data points that are not explained by the model or we use large amounts of observed data. So, instead of maximizing the likelihood function, we minimize the $L^1$ distance between the empirical density function (histogram) $p_e(x,t)$ ($p_e(x,h)$) and the estimated density function $p(x,t;h|r,K,\sigma)$. For the sake of simplicity, we consider the estimate of parameters of the separate function $p_1(x,t|r_1,K_1,\sigma_1)$ ( $p_2(x,h|r_2,K_2,\sigma_2)$ ). Empirical density function $p_e(x,t)$ ($p_e(x,h)$) depends on the observed data and estimated density function $p_1(x,t|r_1,K_1,\sigma_1)$ ( $p_2(x,h|r_2,K_2,\sigma_2)$ ) depends on the used stochastic growth law. The $L^1$ distance has the following form (Rupšys 2005)

$$d_1(r_1,K_1,\sigma_1) = \frac{1}{m}\sum_{j=1}^{m}\int_0^{+\infty}\left|p_e(x,t^j) - p_1(x,t^j|r_1,K_1,\sigma_1)\right|dx \qquad (13)$$

$$d_2(r_2,K_2,\sigma_2) = \frac{1}{m}\sum_{j=1}^{m}\int_0^{+\infty}\left|p_e(x,h^j) - p_2(x,h^j|r_2,K_2,\sigma_2)\right|dx \qquad (14)$$

where $m$ is the number of steps, $t^j \in [0;T_{\max}]$, ( $h^j \in [0;H_{\max}]$),*j=1,2,...m.* In order to simulate numerically the integral defined by the right-hand side of equations (13)-(14), we define the empirical density as

$$p_e(x^i,t^j) = \frac{1}{\Delta\cdot n}\sum_{k=1}^{n} 1_{\{x^i-\Delta/2\le x_k<x^i+\Delta/2\}\{t^j-\Delta1/2\le t_k<t^j+\Delta1/2\}},\ j=1,2,...m.$$

$$p_e(x^i,h^j) = \frac{1}{\Delta\cdot n}\sum_{k=1}^{n} 1_{\{x^i-\Delta/2\le x_k<x^i+\Delta/2\}\{h^j-\Delta1/2\le h_k<h^j+\Delta1/2\}},\ j=1,2,...m.$$

where $1_{\{x^i-\Delta/2\le x_k<x^i+\Delta/2\}\{t^j-\Delta1/2\le t_k<t^j+\Delta1/2\}}$ is one if the observation $x_k$ is in $\left[x^i-\Delta/2;x^i-\Delta/2\right[$, observation $t_k$ is in $\left[t^j-\Delta1/2;t^j-\Delta1/2\right[$ and zero otherwise, $x^i=\Delta\cdot i$ , $n$ is the number of observations, $\Delta$ is the step size, $t^j = \Delta1\cdot j$, $\Delta1$ is the step size ( $1_{\{x^i-\Delta/2\le x_k<x^i+\Delta/2\}\{h^j-\Delta1/2\le h_k<h^j+\Delta1/2\}}$ is one if the observation $x_k$ is in $\left[x^i-\Delta/2;x^i-\Delta/2\right[$, the observation $h_k$ is in $\left[h^j-\Delta1/2;h^j-\Delta1/2\right[$ and zero otherwise, $h^j = \Delta1\cdot j$, $\Delta1$ is the step size). Hence, the numerical approximation of equation (13)-(14) takes the form

$$d_1(r_1,K_1,\sigma_1) \approx \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{\infty}\left|p_e(x^i,t^j) - p_1(x^i,t^j|r_1,K_1,\sigma_1)\right|\Delta \qquad (15)$$

$$d_2(r_2,K_2,\sigma_2) \approx \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{\infty}\left|p_e(x^i,h^j) - p_2(x^i,h^j|r_2,K_2,\sigma_2)\right|\Delta \qquad (16)$$

Our interest centers on the minimizing of the $L^1$ distance function defined by (15)-(16). Hence $\min_{r_k, K_k, \sigma_k} d_k(r_k, K_k, \sigma_k)$

In this study derived mixture (10) of probability density functions is compared to the commonly used Weibull distribution with probability density function (1) and the negative exponential distribution with probability density function

$$p(x) = \begin{cases} a\exp(bx), x \geq 0, \\ 0, x < 0 \end{cases}$$

In addition, we use the Chi-squared test to evaluate the goodness-of-fit of the estimated probability density functions to the observations of diameter frequency distribution. If the calculated p-value of the test is more than a significance level (0.01), then we cannot reject the null hypothesis that the estimated distribution is coincident with the empirical distribution. Conversely, if the Chi-squared test rejected the null hypothesis, it merely means that the estimated tree diameter distribution and the width of the diameter classes were not perfectly selected. Consequently, an error probability (the significance level) is chosen 0.01.

The Chi-squared test is normally applied for the univariate distributions. Our Gompertz type probability density functions (6)-(7) of tree diameter distribution depend on the age and height. So far, the only way to carry out the Chi-squared test is to describe the diameter distribution by a set of overlapping the Gompertz type distributions, which depend on separate age and height classes. The overall diameter distribution of tree in a stand is expressed as the mixture of diameter distributions at different definite ages (heights). The mixture of probability density functions of tree diameter could be described in the following form

$$p(x) = \alpha_1 \sum_{j=1}^{m_1} \lambda_j p(x, t_j) + \alpha_2 \sum_{j=1}^{m_2} w_j p(x, h_j) \qquad (17)$$

$$\sum_{j=1}^{m_1} \lambda_j = 1, \ \sum_{j=1}^{m_2} w_j = 1, \ \alpha_1 + \alpha_2 = 1,$$

where $m_1$ is the number of groups according to the age classes, $m_2$ is the number of groups according to the height classes, $\lambda_j$ is the part of the stand with $j$ age trees, $w_j$ is the part of the stand with height trees. The weights $\alpha_1, \alpha_2 = 1/2$, $\lambda_j$, $w_j$ are calculated using the observed data set and presented in Table 1. The mean values of age and height class are presented in Table 1 too.

Considering the limitations of the Chi-squared test, we calculate three other type of the goodness of fit statistics, namely, the absolute discrepancy, the Kolmogorov-Smirnov, and the Reynolds error index.

**Table 1.** Weights and mean values used in the Gompertz type density mixture

| Class | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 0.4352 | 0.0556 | 0.0860 | 0.0873 | 0.0961 | 0.1107 | 0.0721 | 0.0405 | 0.0114 | 0.0051 |
| 16.0203 | 25.7841 | 34.4559 | 44.1087 | 54.3355 | 64.2229 | 74.0000 | 83.7031 | 94.0556 | 101.6250 |
| 0.0765 | 0.3618 | 0.0500 | 0.0949 | 0.1075 | 0.1278 | 0.1290 | 0.0474 | 0.0051 | |
| 3.3248 | 5.4598 | 9.7911 | 14.0487 | 17.8288 | 21.8005 | 25.7338 | 29.0360 | 32.7500 | |

The absolute discrepancy ($AD$) measures the difference between empirical distribution and the estimated distribution and has the following form

$$AD = \frac{1}{2} \sum_{i=1}^{m} \left| \hat{f}_i - f_i \right|$$

where $m$ is the number of diameter classes, $\hat{f}_i, f_i$ are the relative frequencies of trees in diameter classes of the estimated and empirical distributions. The estimated and empirical distributions have nothing in common if $AD=1$, they are identical if $AD=0$.

The Kolmogorov-Smirnov statistics ($KS$) has the following form

$$KS = \max\left\{ \max_{1 \leq j \leq n} \left[ \frac{j}{n} - u_j \right], \max_{1 \leq j \leq n} \left[ u_j - \frac{j-1}{n} \right] \right\}$$

where $n$ is the number of observations, $u_j = F(x_{(j)})$, $F$ is the estimated cumulative distribution function, and the $x_{(j)}$ are observations sorted in ascending order for diameter.

The Reynolds error index ($EI$) has the following form

$$EI = \sum_{i=1}^{m} \left| \hat{n}_i - n_i \right|$$

where $\hat{n}_i$, $n_i$ are the estimated and empirical number of trees in diameter class $i$. The Reynolds error index is calculated in 5 cm classes using the number of sample trees as weight.

Next we use the so-called pseudo-residuals defined by Zucchini and MacDonald (1999) as an alternative method for the assessment of the goodness-of-fit. The key assumptions are that observations $X_1$, $X_2$, ..., $X_n$ are independent and have distribution function $F_i(x)$ (note, that $X_i$ are not assumed to be identically distributed). The pseudo-residual, $r_i$, corresponding to observation $x_i$ is defined as

$$r_i = \Phi^{-1}(F_i(x_i)), \ i = 1, 2, ..., n,$$

where $\Phi^{-1}$ denotes the inverse function of the distribution function of the standard normal distribution.

It is shown (Zucchini *et al*. 2001) that pseudo-residuals $r_i$, $i = 1,2,...,n$ follow the standard normal distribution if estimated density function $f_i(x)$ is indeed the correctly specified density function for the observations $X_1$, $X_2$, ..., $X_n$. A basic graphical approach for checking normality of the pseudo-residuals $r_i$, $i = 1,2,...,n$ is the normal quantile plot (Wilks *et al.* 1968, Looney *et al.* 1984). The normal quantile plot compares the *i*th ordered value $r_{(i)}$ with the *i/(n+1)*th quantile (rankits) of the standard normal distribution defined by

$$q_i = \Phi^{-1}\left(\frac{i}{n+1}\right),\ i = 1,2,...,n\ .$$

More alternative types of plotting positions we can see (Gan *et al.* 1991). A sample from the standard normal distribution will result in straight line *r=q* on a normal quantile plot. Any deviation from this line will indicate lack of fit of the estimated probability density function of diameter distribution. A normal quantile plot is sensitive in detecting distinctions in the tail region of tree diameter distribution. In our study, the curvature of a normal quantile plot is exploited to validate or reject the estimated tree diameter distribution. To assess significance of departures from linearity of the normal quantile plot, we additionally calculate 1% upper and lower simulation envelopes (Diggle 1983) and plot against $q_i$, $i = 1,2,...,n$ .

## Results and discussion

The all used probability density functions are fitted to the observations described in Figure 1 using the $L^1$ distance procedure. The width of the diameter classes is chosen approximately 5 cm. The $L^1$ distance functions (15)-(16) are minimizing for the estimation of parameters of the probability density functions defined by (1), (6), (7), (17). The resulting $L^1$ distance functions are nonlinear in parameters and so cannot be solved exactly. For the calculation of estimates, a number of iterative procedure can be used. It is known that the iterative procedures may converge very slowly, or oscillate widely, or may not converge at all. So, we use a Monte Carlo approach. The results are presented in Table 2.

It will be observed that the $L^1$ distance values obtained in predicting parameters for all used distributions are very similar (see Table 2).

Figure 2 shows empirical probability density function (histogram) $p_e(x,t^j)$ ( $p_e(x,h^j)$ ) *j=1,2…,4* and estimated probability density function $p_1(x,t^j|r_1,K_1,\sigma_1)$ ( $p_2(x,h^j|r_2,K_2,\sigma_2)$) *j=1,2…,4* for the given four ages 20, 40, 60, and 80 years (four heights 5, 15, 25, and 35 metres). The fit is not always so good. Indeed, for the age $L^1$ distance is 0.3402.

**Table 2.** Parameter estimates for predictive diameter distributions

| Study case | Parameters of the Gompertz type density | | | $L^1$ norm |
|---|---|---|---|---|
| | *r* | *K* | *σ* | |
| Age (6) | 0.0331 | 45.5459 | 0.0794 | 0.3402 |
| Height (7) | 0.0793 | 51.2156 | 0.1037 | 0.2818 |
| | Parameters of the Weibull and negative exponential densities | | | $L^1$ norm |
| | *a* | *b* | *c* | |
| Weibull (1) | 0.9380 | 21.5020 | 1.3887 | 0.3119 |
| Exponential (17) | 0.0469 | -0.0545 | | 0.3845 |

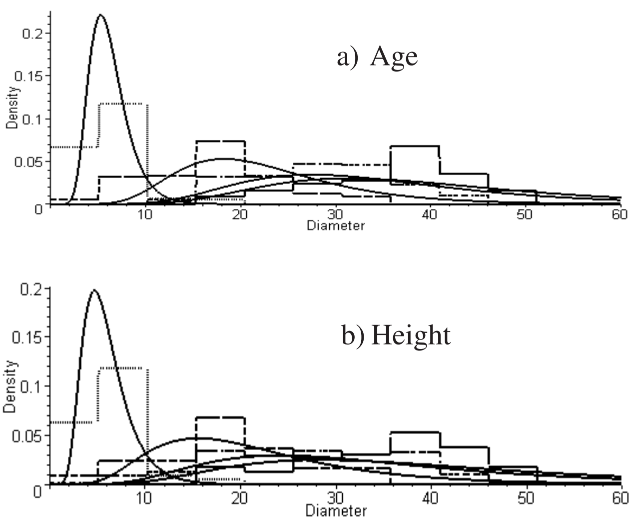



**Figure 2.** Plot of the empirical and estimated Gompertz type probability density functions: a) the ages 20, 40, 60, and 80 years, b) the heights 5, 15, 25, and 35 metres

By examining Figure 2 it is clear that the variation of tree diameter is low for small values of age (height) and increases for higher values of age (height). To show this variation, the probability density functions $p_1(x,t|r_1,K_1,\sigma_1)$, $p_2(x,h|r_2,K_2,\sigma_2)$ are plotted in Figure 3.

To provide equations for diameter growth, two different dynamic diameter growth age dependent and height dependent models (2)-(3) are differentiated in this study. Using the estimates of parameters presented in Table 2, we calculate a numerical approximation of the solutions of the system equations (8)-(9), which represent mean diameter growth trend and variance. The numerical approximations of mean trend and the

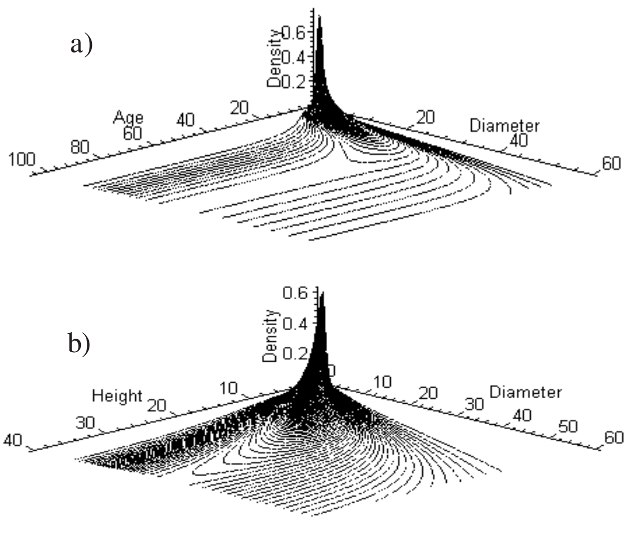



**Figure 3.** Plot of the Gompertz type probability density functions: a) the age dependent $p_1(x,t|r_1,K_1,\sigma_1)$, b) the height dependent $p_2(x,h|r_2,K_2,\sigma_2)$

standard deviation of tree diameter are plotted in Figure 4. For both models, the estimated diameter trajectories are very close. The above illustrations in Figures 2, 3, 4 exemplify some interesting kinetic features of diameter distribution, when the age and height are predictors. The mean trend of diameter evolves monotonically toward the value of the diameter carrying capacity $K$. The diameter variance over age (height) grows monotonically towards the steady state value.

Figure 5 shows the empirical and estimated (mixture (18), Weibull (1) and negative exponential (17)) diameter probability density functions as an example of a preference of the Gompertz type mixture (18). The

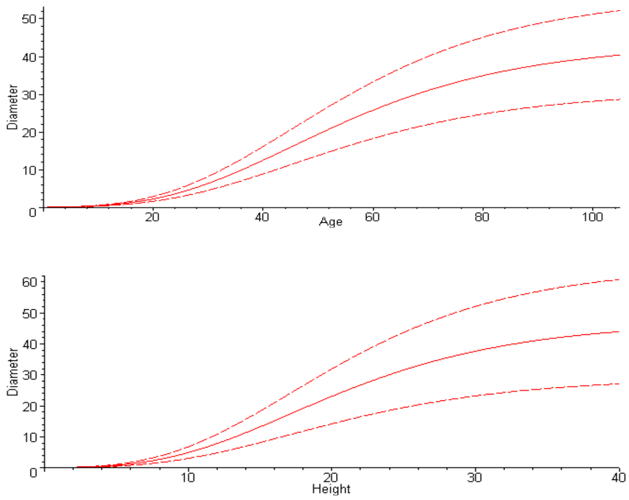



**Figure 4.** Plot of the diameter growth (mean trend - solid line, mean trend ± standard deviation – dash line): a) the age dependent, b) the height dependent

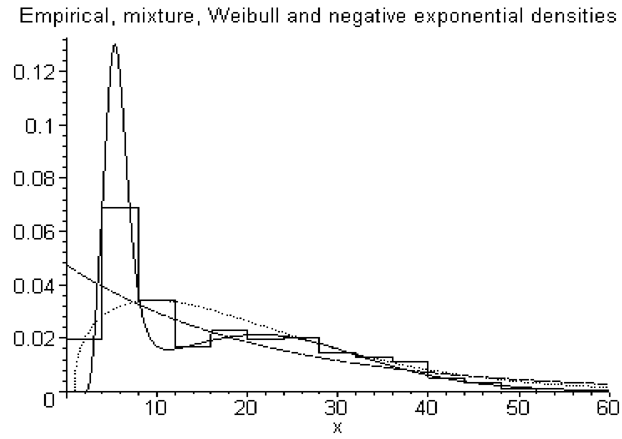



**Figure 5.** Empirical and estimated diameter distributions: the Gompertz type mixture (18) - solid line, the Weibull (1) – dot line, the negative exponential (17) – dash line, the empirical – piecewise linear

age and height dependent mixtures (10), (18) produce the diameter distribution at various stages of stand development and provide forest managers with a tool for assessment of stands.

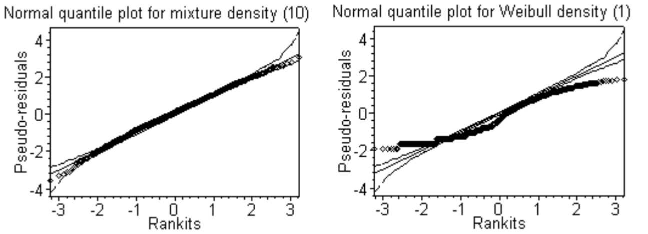
The summaries of the goodness-of-fit of the absolute discrepancy, Kolmogorov-Smirnov and Reynolds statistics are presented in Table 3.

**Table 3.** Goodness-of-fit statistics and shi-square test

| Density | Statistics | | | $\Pr ob(X > \chi^2)$ ($\chi^2$) |
|---|---|---|---|---|
| | *AD* | *KS* | *EI* | |
| Gompertz type (18) | 0.0393 | 0.0705 | 124 | 0.0387 (14.7956) |
| Weibull (1) | 0.1727 | 0.1911 | 547 | <0.0001 (229.3710) |
| Exponential (17) | 0.1738 | 0.1101 | 549 | <0.0001 (166.3869) |

An analysis of the results in Table 3 indicates that *AD* and *EI* tests ranked all used distributions very similar. Our derived mixture (18) of ranked best among the used distributions and the negative exponential was the poorest performer. In comparison to the Weibull distribution (1), which successfully describes unimodal diameter distributions, this mixture reduced the *AD* statistic by 77%, the *KS* statistic by 63%, and the *EI* statistic by 77%. We use the Chi-squared test to test the hypothesis that the empirical and estimated distributions are statistically identical, with significance 1%. Verifying the null hypothesis that the diameter distribution of tree has the estimated form (mixture (18), Weibull, negative exponential) against the alternative hypothesis that this distribution differs from estimated was calculated p-value. For the mixture (18), the p-value 0.0387 indicates a large amount of evidence that this mixture distribution has good performance.

The normal quantile plots of pseudo residuals of all estimated diameter distributions together with their envelopes are shown in Figure 6.



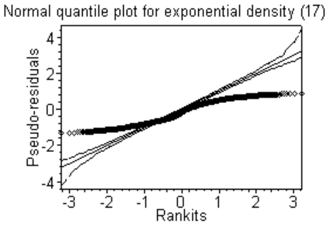




**Figure 6.** Normal quantile plots of the pseudo residuals for all estimated distributions: the normal quantile plot - solid line, the envelopes – dot line, the linear $r=q$ – dash line

In this paper, the curvature of a normal quantile plot is exploited to identify a suitable theoretical density to fit a data set. Just like other graphical methods, the procedure proposed here is an inaccurate procedure. From Figure 6, it is observed that the normal quantile plot of the Weibull and negative exponential densities are nonlinear and corresponds well with the mixture density (10). In an attempt to quantify the straightness of the normal quantile plot, Shapiro and Francia (1972) proposed summary statistics $W'$, which is the squared correlation between the pairs of point in the plots. So, normal pseudo-residuals should display a correlation coefficient near 1 and, thus, the square of the correlation coefficient, $W'$, should be near 1. If $W'$ is too small, it indicates that the pseudo-residuals are inconsistent with the requirement of normality. The pseudo-residuals calculated for the mixture density (10) yield a $W'$ value 0.9971, for the Weibull density (1) yield 0.9537, and for the negative exponential density (17) yield 0.9491. These values should be compared to $W(0.01;1581)$=0.9962. Of these three values two first are considerably below the 1% critical point. The normal quantile plot and the $W'$ value of the pseudo-residuals for the Gompertz type mixture (10) show that the plot is much more linear and indicate the satisfactory fit of the Gompertz type mixture (10) to the observations.

## Conclusions

A wide variety of stand growth kinetics laws lead to different tree diameter distribution. Interpretations of the tree diameter distribution with stand age and height structure minimizes the risk to be in the wrong choice of theoretical distribution. The main goal of this study was to introduce a new method for the modeling of diameter distribution of tree, demonstrate its use, show how the proposed method works to the field data set, and compare with two commonly used probability density functions of tree diameter distribution – the Weibull distribution, and the negative exponential distribution. It was found that the Gompertz type mixture of probability density functions fits the observed data very well. The proposed method can be used in practical forestry applications.

In the final note, we wish to point out that the method implemented here for obtaining the diameter distribution, based on postulated total diameter growth kinetics as a function of the age (height), can be modified incorporating more growth laws. For example, we can assume that the Verhulst, Bertalanffy, Mitcherlich, Richards (or their modifications) laws describe diameter growth curves best.

The mixture of probability density functions of tree diameter distribution can be expanded for other predictor stand-level variables, for example, the number of trees per ha, the relative spacing, the site occupancy, etc.

To cover more dynamic stochastic diameter growth laws, the diffusion term can be changed too.

In order to propose more precise dynamic stochastic growth law, the stochastic delay differential equations can be used.

## References

**Alvarez, J. G., Schroder, J., Rodriguez, R. and Ruiz, A. D.** 2002. Modelling the effect of thinnings on the diameter distribution of even-aged maritime pine stands. *For. Ecol. Manage.*, 165: 57-65.

**Ananda, M. M., Dalpatadu, R. J. and Singh, A. K.** 1996. Adaptive Bayes estimators for parameters of the Gompertz survival model. *Applied Mathematics and Computation,* 75: 167-177.

**Bailey, R. L. and Dell, T. R.** 1973. Qualifying diameter distributions with the Weibull function. *For. Sci.*, 19: 97–104.

**Bliss, C. I and Reinher, K. A.** 1964. A log–normal approach to diameter distribution in even–aged stands. *For. Sci.*, 10, 350 – 360.

**Bruchwald, A.** 1988. Parameters of the distribution of B. H. diameters of trees in pine stands. Ann. Warsaw Agric. Univ. – SGGW – AR, F*or. Wood technol.*, 37: 75 – 83.

**Cao, V.** 2004. Predicting parameters of a Weibull function for modeling diameter distribution. *For. Sci.*, 50: 682 – 685.

**Chen, W.** 2004. Tree size distribution functions of four borel forest types for biomass mapping. *For. Sci.,* 50: 436 – 449.

**Clutter, J. L. and Bennett, F. A.** 1965. Diameter distributions in old–field slash pine plantations. GA *For. Res. Counc. Rep.*, 13: 9.

**De Liocourt, F.** 1898. De l'amenagement des sapinieres. Bulletin trimestriel 4, *Societe Forestiere de Franche-Compte et Belfort*, 645 p.

**Diggle, P. J.** 1983. Statistical analysis of spatial point patterns. Academic Press, New York, 148 p.

**Dieguez-Aranda, U., Castedo Dorado, F., Alvarez Gonzalez, J. G. and Rojo Alboreca, A.** 2005. Dynamic growth model for Scots pine (*Pinus sylvestris* L.) plantations in Galicia (north-western Spain). *Ecol. model.*, 191: 225-242.

**Gan, F. F., Koehler, K.J. and Thompson, J.P.** 1991. Probability plots and distribution curves for assessing the fit of probability models. *Amer. Statist.*, 45: 14-21.

**Garcia, O.** 1983. A stochastic differential equation model for the height growth of forest stands. *Biometrics*, 39: 1059-1072.

**Garcia, O.** 1994. The state–space approach in growth modelling. *Can. J. For. Res.*, 24: 1894 – 1903.

**Garcia, O.** 2005. Comparing and combining stem analysis and permanent sample plot data in site index models. *For. Sci.*, 51: 277-283.

**Гихман, И.И. and Скороход, А.В.** 1977. Введение в теорию случайных процессов [Introduction to theory of random processes]. *Наука*, Москва, 567 p. (in Russian)

**Gompertz, B.** 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies, *Phil. Trans. Roy. Soc.*, 115: 513-585.

**Hafley, W. L. and Shreuder, H. T.** 1977. Statistical distributions for fitting diameter and height data in even–aged stands. *Can. J. For. Res.*, 7: 481 – 487.

**Kangas, A. and Maltamo, M.** 2000. Calibrated predicted diameter distribution with additional information. *For. Sci.*, 46: 390-396 (2000).

**Knoebel, B. R. and Burkhart, H. E.** 1991 A bivariate distribution approach to modeling forest diameter distributions at two points of time. *Biometrics,* 47: 241-253.

Kubilius, J. 1980. Tikimybių teorija ir matematinė statistika [Probability theory and mathematical statistics]. *Mokslas*, Vilnius, 468 p. (in Lithuanian).

**Leak, W. B.** 1964. An expression of diameter distribution for unbalanced, uneven-aged stands and forests. *For. Sci.*, 10: 39-50.

**Leak, W. A.** 1996. Long-term structural change in uneven-aged northern hardwoods. *For. Sci.*, 42: 160-165.

**Li, F., Zhang, L. and Davis, C. J.** 2002. Modeling the joint distribution of tree diameters and heights by bivariate generalized beta distribution. *For. Sci.*, 48: 47-58.

**Lindsay, S. R., Wood, G. R. and Woollons, R. C.** 1996. Stand table modeling throught the Weibull distribution and usage of skewness information. *For. Ecol. Manage.*, 81: 19-23.

**Looney, S. W. and Gulledge, Jr. T. R.** 1984. Probability plotting positions and goodness of fit for normal distribution. *The Statistician*, 34: 297-303.

**Maltamo, M., Kangas, A., Uuttera, J., Torniainen, T. and Saramaki J.** 2000. Comparison of percentile based prediction methods and the Weibull distribution in describing the diameter distribution of heterogeneous Scots pine stands. *For. Ecol. Manage.*, 133: 263-274.

**Matis, J. H., Kiffe, T. R., Renshaw, E. and Hassan, J.** 2003. A simple saddlepoint approximation for the equilibrium distribution of the stochastic logistic model of population growth. *Ecol. Model.*, 161: 139-248.

**McCarthy, J. W.and Weetman, G.** 2006. Age and size structure of gap-dynamic, old-growth borel forests stands in Newfoundland. *Silva Fennica*, 40: 209-230.

**Mehtatalo, L.** 2005. Localizing a predicted diameter distribution using sample information. *For. Sci.*, 51: 292-303.

**Meyer, H. A. and Stevenson, D. D.** 1943. The structure and growth of virgin beech, birch, maple, hemlock forests in northern Pensylvania. *J. Agric. Res.*, 67: 465-478.

**Meyer, H. A.** 1952. Structure, growth, and drain in balanced, uneven-aged forests. *Journal of Forestry*, 50: 85-92.

**Nelson, T. C.** 1964. Diameter distribution and growth of loblolly pine. *Forest Sci.*, 10: 105 – 115.

**Newton, P. F.** 1997. Stand density management diagram: review of their development and utility in stand-level management planning. *For. Ecol. Manage.*, 98: 251-265.

**Newton, P. F., Lei, Y. and Zhang, S. Y.** 2005. Stand-level diameter distribution yield model for black spruce plantations. *For. Ecol. Manage.*, 209: 181-192.

**Nikulin, M. S.** 1973. The Chi-squared test for continuous distributions. *Prob. Theory and Their Appl.* 18(3): 159-172.

**Nord-Larsen, T. and Cao, Q. V.** 2006. A diameter distribution model for even-aged beech in Denmark. *For. Ecol. Manage.*, 231: 218-225.

**Prodan, M.** 1953. Die Verteilung des Vorrates gleichaltriger Hochwaldbestande auf Durchmesserstufen. *Allgemeine Forst – und Jagdzeitung*, 124: 93-106.

**Reynolds, M. R., Burk, T. E. and Huang, W.** 1988. Goodness-of-fit tests and model selection procedures for diameter distribution functions. *For. Sci.*, 34: 373-399.

**Rupšys, P.** 2003. Stochastic analysis of logistic growth models. In: U. Olsson, J. Sikk (Editorss), Fourth Nordic – Baltic Agrometrics Conference. Proceedings of the International Conference, *Swedish University of Agricultural Sciences, Department of Biometry and Informatics, Report 81*, Uppsala, Sweden, 127-139.

**Rupšys, P.** 2004. On parameter estimation for stochastic logistic growth laws through the maximum likelihood procedure. *Lith. Math. J.*, 44(spec. Nr.): 759-762.

**Rupšys, P.** 2005. On parameter estimation for stochastic logistic growth laws through the maximum likelihood and $L^1$ procedure. *Lithuanian Statistics: Articles, Reports and Studies*, 42: 49-60.

**Schnur, G. L.** 1934. Diameter distribution for old–field loblolly pine stands in Maryland. *J. Agric. Res.*, 49: 731 – 743.

**Shapiro, S. S. and Francia, R. S.** 1972. An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67: 215-216.

**Shoi, I.** 1998. Approximation of continuous time stochastic processes by a local linearization method. *Math. Comput.*, 67: 287-298.

**Shoi, I. and Ozaki, T.** 1997. Comparative study of estimation methods for continuous time stochastic processes. *J. Time Ser. Anal.:* 18; 485-506.

**Soboleva, T. K. and Pleasants, A. B.** 2003. Population growth as a nonlinear stochastic process. *Mathematical and Computer Modelling*, 38: 1437-1442.

**Sterba, H.** 2004. Equilibrium curves and growth models to deal with forests in transition to uneven-age structure – application in two sample stands. *Silva Fennica*, 38: 413-423.

**Tang, S., Wang, Y., Zhang, L. and Meng, C. H.** 1997. A distribution-independent approach to predicting stand diameter distribution. *For. Sci.*, 43: 491-500.

**Tewari, V. P. and von Gadow, K.** 1999. Modelling the relationships between tree diameters and heights using $S_{BB}$ distribution. *Forest Ecol. and Manage.*, 119: 171-176.

**Tham, A.** (1988). Structure of mixed *Picea abbies* (L.) Karst. and *Betula pendula* Roth. and *Betula pubescens* Ehrh. stands in south and middle Sweden. *Scand. J. Forest Res.*, 3: 355-370.

**Uuttera, J. and Maltamo, M.** 1995. Impact of regeneration method on stand structure prior to first thinning. Comparative study, North Karelia, Finland vs. Republic of Karelia, Russian Federation. *Silva Fennica*, 29: 267-285.

**Uusitalo, J. and Kivinen, V.P.** 1998. Constructing bivariate dbh/dead-branch height distribution of pines for use in

sawing production planning. *Scandinavian Journal of Forest Research*, 13: 509-514.
**Weibull, W.** 1951. A statistical distribution function of applicability. *J. Appl. Mech*., 18: 293-297.
**Westphal, C., Tremer, N., von Oheimb, G., Hanse, J., von Gadow, K. and Hardtle, W.** 2006. Is a reverse J-shaped diameter distribution universally applicable in European virgin beech forests? *For. Ecol. Manage.*, 223: 75-83.
**Wilks, M. B. and Gnanadesikan, R.** 1968. Probability plotting methods for the analysis of data. *Biometrika*, 55: 1-19.
**Zasada, M. and Ciesziewski, C. J.** 2005. A finite mixture distribution approach for characterizing tree diameter distributions by natural social class in pure even–aged Scots pine stands in Poland. *For. Ecol. Manage.*, 204: 145 – 158.
**Zucchini, W. and MacDonald I.L.** 1999. Illustrations of the use of pseudo-residuals in assessing the fit of a model. In: A. Berhold, G. Kauerman (Editors), Proceedings of the 14th International Workshop of Statistical Modelling, 19-23 July 1999, Graz. Friedl, 409-416.
**Zucchini W., Schmidt M. and von Gadow K.** 2001. A model for the diameter-height distribution in an uneven-aged beech forest and a method to assess the fit of such models. *Silva Fennica*, 35: 169-182.

Received 20 March 2007
Accepted 30 October 2007

## СТОХАСТИЧЕСКИЙ ЗАКОН РОСТА ТИПА ГОМПЕРТЦА И РАСПРЕДЕЛЕНИЕ ДИАМЕТРА ДЕРЕВЬЕВ

**П. Рупшис, Э. Пятраускас, Ю. Мажейка и Р. Дялтувас**

*Резюме*

В статье рассматривается вопросы распределения деревьев по диаметру. В основу разработки модели распределения положен вероятностный логистический закон роста. Используя стохастический закон роста Гомпертца разработана функция распределения деревьев по диаметру в зависимости от возраста и средней высоты древостоя. В статье приводятся сравнения разработанной модели с распределением Вейбулла и отрицательным экспоненциальным. Доказано преимущество разработанной модели. Моделирование распределения диаметра осуществлено на реальных данных, полученных на пробных площадях в Дубравском лесхозе. Разработанную модель распределения деревьев по диаметру целесообразно использовать для моделирования роста  инвентаризации лесов. В работе использован язык MAPLE.

**Ключевые слова:** Вейбулл, Гомпертц, распределение диаметра, стохастический закон роста, уравнение Фоккер-Планса