

SCIENTIFIC REPORTS



OPEN

A new paradigm in modelling the evolution of a stand via the distribution of tree sizes

Petras Rupšys^{1,2} & Edmundas Petrauskas¹

Our study focuses on investigating a modern modelling paradigm, a bivariate stochastic process, that allows us to link individual tree variables with growth and yield stand attributes. In this paper, our aim is to introduce the mathematics of mixed effect parameters in a bivariate stochastic differential equation and to describe how such a model can be used to aid our understanding of the bivariate height and diameter distribution in a stand using a large dataset provided by the Lithuanian National Forest Inventory (LNFI). We examine tree height and diameter evolution with a Vasicek-type bivariate stochastic differential equation and mixed effect parameters. It is focused on demonstrating how new developed bivariate conditional probability density functions allowed us to calculate the evolution, in the forward and backward directions, of the mean diameter, height, dominant height, assortments, stem volume of a stand and uncertainties in these attributes for a given stand age. We estimate the parameters by considering discrete samples of the diameter and height at a given age and by using an approximated maximum likelihood procedure. The model performance criteria for the height and diameter growth models include statistical indexes and an analysis of residuals.

Foresters seek to predict patterns in the distributions of individual trees within stands, across tree sizes and over space. Determining a distribution of tree diameters and heights has been broadly studied since the twentieth century¹. Originally, the tree diameter and height distribution concept crystallised in assessing standing timber value based on merchantable piece sizes and forest structure which are the basis of aboveground forest carbon estimates². The distribution of tree diameters at the breast height (in the sequel – diameter) is one of the most investigated subjects of statistical research in forestry. The forestry literature reports that tree diameter distributions vary across different stands³. In forest research, the univariate tree diameter distribution was introduced by Meyer⁴ and Schnur⁵ using a discrete probability space methodology. Natural extensions of a discrete diameter distribution are the continuous univariate gamma distribution⁶, log-normal distribution⁷, Weibull distribution⁸, logit-logistic⁹ among others. Classical mathematical modelling of tree size distribution is largely concerned with the use of a well-known probability density function for fitting to the empirical tree size frequencies and does not quantify the evolution of a tree size distribution via a forest stand age. Several authors have analysed the tree height distribution in a forest stand. The tree height distribution can be depicted using two mathematical models. First, a height distribution can be derived indirectly from a diameter distribution together with the relationship between tree height and diameter. The relationship between height and diameter distributions can be accounted, provided that the function relating to height and diameter can be inverted¹⁰. Second, theoretical height distributions were fitted directly to tree height observations and have included such shapes as the beta¹¹, the Weibull⁸, the Johnson¹², the power-normal¹³ and much more. The construction and applications of bivariate distributions of diameter and height are the most used classical fields of research in forest statistics, and it continues to be an active field of research in forest mensuration and inventory, where both the diameter and height are measured on every tree.

The philosophy of the mean stem volume estimation is based on the modelling of the distributions of the volume components of diameter and height instead of the volume itself. The bivariate distribution, as a possible relevant model for mean stem volume statistics, provides more physical insights on the validity domain and the quantification of the error made by using such a distribution. A bivariate distribution of diameter and height can

¹Institute of Forest Management and Wood Sciences, Aleksandras Stulginskis University, Studentu 11, Kaunas, LT-53361, Lithuania. ²Centre of Mathematics, Physics and Information Technologies, Aleksandras Stulginskis University, Studentu 11, Kaunas, LT-53361, Lithuania. Correspondence and requests for materials should be addressed to P.R. (email: petras.rupsys@asu.lt)

Data	Number of trees	Min	Max	Mean	St. Dev.	Number of trees	Min	Max	Mean	St. Dev.
Estimation						Validation				
d (cm)	5220	15.10	66.10	27.24	8.63	2692	15.10	74.30	27.45	8.83
h (m)	5220	7.10	37.80	22.07	4.42	2692	7.2	34.80	22.26	4.43
A (yr)	5220	17	221	65.78	23.67	2692	17	162	67.70	23.31

Table 1. Datasets summary statistics.

be developed using several mathematical techniques. First, a theoretical diameter and height distribution can be fitted directly to tree diameter and height observations in forms such as the Johnson^{2,14}, the power-normal¹⁵ and others. Second, a theoretical bivariate diameter and height distribution as an extension of univariate marginal distributions can be developed by Plackett's method^{16,17}, the copula method¹⁸ and references therein.

A new paradigm in stand structure, modelling beyond well-defined theoretical distributions, is to use stochastic differential equations (SDEs) techniques. On the other hand, an algebraic structure of SDEs can relate tree size distribution and the age of a forest stand. Nowadays, SDEs find applications in many disciplines including physics¹⁹, information²⁰, biomedicine²¹ and forestry²². The framework of SDEs can also be adapted to modelling theoretical diameter and height distributions²³. One motivation for using SDEs for the modelling of the evolution of the tree size distribution is that unknown drivers functioning in a stand can be modelled as random processes, which lead to a white noise process. The solution to the SDE of the diameter and height dynamics is a bivariate random process which leads to a theoretical bivariate probability density function for the diameter and height. In this philosophy, the evolution of the probability density function for diameter and height, which behaves according to a stochastic differential equation, can be defined by the Fokker-Planck equation. Some models describing the evolution of the fixed effect parameters univariate and bivariate probability density functions were discussed by Rupšys and Petrauskas^{22,23}. Recently, there were also a number of results on mixed effect parameters using the Ornstein-Uhlenbeck family of univariate SDE models of tree height evolution via diameter. Such models generate the normal, or log-normal, probability density functions for diameter and height in a univariate sense^{24–27} or using the normal copula technique in a bivariate sense²⁸.

The newly developed model based on bivariate mixed effects SDE extend the usual non-linear mixed effects regression models through the inclusion of a system noise as an additional source of variation in the first-stage model. This extended model describes the variation in the data through two sources of noise: the system noise, σ , which reflects the random fluctuations around the corresponding theoretical diameter and height models and represents the within-stand variation, and random effect, ϕ , which represents the between-stand variation in the theoretical diameter and height models. If the magnitude of the parameter of the system noise σ is zero, the entire system noise term will vanish, and the remaining part of the SDE will simply be the differential form, the solution of which is the regression term of the mixed effects model. Random effect, ϕ , is not parameter but it is a random variable. The distribution of random effect captures random variation of the parameter in the population of stands and is frequently assumed to be normal.

Nonlinear mixed effect regression models were introduced into forestry management to analyse data from several stands simultaneously. The individual tree size variables are assumed to be described by a common structural model with some of the model parameters varying between the stands (so-called random effects), while other parameters are invariant between stands (so-called fixed effect parameters). The mixed effects methodology can be of great benefit when the model uses the data of the National Forest Inventory as it is regional and sparse.

In this study, the ultimate objective is to build in some detail the bivariate Vasicek type mixed effect parameters SDEs for the diameter and height at a given stand age to deduce the evolution, via stand age, of the bivariate conditional probability density function. This can also be used for predicting stand characteristics (mean tree diameter, height, stem volume, stand volume per ha) and their precisions. The technique for estimating mixed effect parameters is based on an approximation of the maximum log-likelihood function. This work aims to present principal concepts of the newly developed bivariate mixed effect parameters conditional probability density function of the diameter and height in simple terms and to illustrate results using large datasets provided by the Lithuanian National Forest Inventory. All results are implemented in the symbolic algebra system MAPLE.

Materials and Methods

Data. The data used for the newly developed model estimation and validation were obtained from the Lithuanian National Forest Inventory (LNFI) (2007–2010). At plot establishment, the following data were recorded for every sample tree: the species, the age, the diameter over bark at 1.30 m height and measured to the nearest millimetre. The total tree height was measured for a subsample of trees (no less than 3) to the nearest quarter metre. The tree diameters were measured with field callipers in two perpendicular directions. In the main 500 m² size circular plot (horizontal radius equals to 12.62 m) all larger than 14.0 cm in diameter trees at 1.3 m height were inventoried. A total of 2,108 plots of Scots pine trees were chosen from the LNFI 2007–2010 database. The dataset was randomly divided into estimation and validation datasets. A random sample of 1,408 plots (5,220 trees) was selected for model estimation, and the remaining dataset of 700 plots (2,692 trees) was utilised for model validation. Only measurements from live trees without top damage were included in the statistical analysis. Summary statistics for the diameter at breast height (d), the total height (h) and the age (A) for all of the trees used in model estimation and validation datasets are presented in Table 1. It should be noted that data on the number of plots with greater than 10 measured trees (diameter and height) were very limited.

Bivariate stochastic differential equation framework. In this paper, we take a bivariate Ornstein-Uhlenbeck family²⁹ SDE of the Vasicek type²² to study the tree diameter and height bivariate distribution problem. This results in an exact bivariate conditional probability density function which parameters can be estimated by maximum likelihood procedure based on discrete time observations. A bivariate mixed effect parameters SDE is set up for describing the tree diameter and height dynamics that extends to a theory of a bivariate conditional probability density function of the tree diameter and height. Proceeding as we have in the fixed effect parameters Gompertz-type bivariate case²³, the Vasicek type bivariate SDE that describes the development of the diameter and height is defined by:

$$dX^i(t) = A(X^i(t))dt + B^{\frac{1}{2}} \cdot dW(t), \quad i = 1, 2, \dots, M, \tag{1}$$

here: M is the total number of stands used for model fitting, t is the time (stand age), $X(t) = (D(t), H(t))^T$, $t \in [t_0, T]$, $t_0 \geq 0$, $X(t_0) = x_0 = (d_0, h_0)^T$, $d_0 \geq 0$, $h_0 \geq 0$, the drift vector $A(x)$ is defined as:

$$A(x) = (\beta_d(\alpha_d + \phi_d^i - d), \beta_h(\alpha_h + \phi_h^i - h))^T, \tag{2}$$

the diffusion matrix B is defined as:

$$B = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}, \tag{3}$$

$W^i(t) = (W_d^i(t), W_h^i(t))^T$, $t \in [t_0, T]$ are independent bivariate Brownian motions, ϕ_d^i and ϕ_h^i , $i = 1, 2, \dots, M$, are independent and normally distributed random variables with zero mean and constant variances ($\phi_d^i \sim N(0; \sigma_d^2)$ and $\phi_h^i \sim N(0; \sigma_h^2)$), $\alpha_d, \alpha_h, \beta_d, \beta_h, \sigma_{11}, \sigma_{12}, \sigma_{22}, \sigma_d, \sigma_h$ are fixed effect parameters to be estimated and $W^i(t)$, $W^j(t)$, $\phi_d^i, \phi_d^j, \phi_h^i$ and ϕ_h^j are mutually independent for all $1 \leq i, j \leq M$. In the sequel, initial condition takes the form: if $t_0 = 5$, then $d_0 = 0.0$ and $h_0 = 1.3$.

The Vasicek-type bivariate SDE can be converted into a well-studied bivariate Ornstein-Uhlenbeck²⁹ process by the transformation $Y(t) = (e^{\beta_d t} \cdot D(t), e^{\beta_h t} \cdot H(t))^T$ and solved explicitly. By applying the Ito formula³⁰ and taking into account that the random vector $(\int_{t_0}^t e^{\beta_d s} \cdot W_d^i(s), \int_{t_0}^t e^{\beta_h s} \cdot W_h^i(s))^T$ has a bivariate normal distribution, we can deduce that the conditional random vector $(X^i(t)|X(t_0) = x_0) = (D(t)|D(t_0) = d_0, H(t)|H(t_0) = h_0)^T$ has a bivariate normal distribution $N_2(\mu^i(t); \Sigma(t))$, $i = 1, 2, \dots, M$, with the mean vector $\mu^i(t) = (\mu_d^i(t), \mu_h^i(t))^T$:

$$\mu_d^i(t) = \alpha_d + \phi_d^i - (d_0 - (\alpha_d + \phi_d^i)) \cdot e^{-\beta_d(t-t_0)}, \tag{4}$$

$$\mu_h^i(t) = \alpha_h + \phi_h^i - (h_0 - (\alpha_h + \phi_h^i)) \cdot e^{-\beta_h(t-t_0)}, \tag{5}$$

the variance-covariance matrix $\Sigma(t)$:

$$\Sigma(t) = \begin{bmatrix} \nu_d^2(t) & \nu_{d,h}(t) \\ \nu_{d,h}(t) & \nu_h^2(t) \end{bmatrix}, \tag{6}$$

$$\nu_d^2(t) = \frac{\sigma_{11}(1 - e^{-2\beta_d(t-t_0)})}{2\beta_d}, \tag{7}$$

$$\nu_h^2(t) = \frac{\sigma_{22}(1 - e^{-2\beta_h(t-t_0)})}{2\beta_h}, \tag{8}$$

$$\nu_{d,h}(t) = \frac{\sigma_{12}(1 - e^{-(\beta_d+\beta_h)(t-t_0)})}{\beta_d + \beta_h}, \tag{9}$$

and conditional probability density function:

$$f(d, h, t | \alpha_d, \alpha_h, \beta_d, \beta_h, \sigma_{11}, \sigma_{12}, \sigma_{22}, \sigma_d, \sigma_h, \phi_d^i, \phi_h^i) = \frac{1}{2\pi dh [\Sigma(t)]^{-1}} \exp\left(-\frac{1}{2} \Omega(d, h, t)\right), \tag{10}$$

$$\Omega(d, h, t) = (d - \mu_d^i(t), h - \mu_h^i(t)) [\Sigma(t)]^{-1} \begin{pmatrix} d - \mu_d^i(t) \\ h - \mu_h^i(t) \end{pmatrix}. \tag{11}$$

Results and Discussion

The conceptually simplest modelling methodology for complex stands evolves from individual-tree level models that are then aggregated to predict consequences at the stand level³¹. Our developed conditional bivariate probability density function, defined by Eq. 10, outlines the evolution of the tree diameter and height structure via a

Parameters										AIC
α_d	β_d	σ_{11}	α_h	β_h	σ_{22}	σ_{12}	σ_d	σ_h		
Univariate fixed effects models										
35.3599 (56*10 ⁻³)	0.0280 (1.1*10 ⁻⁵)	1.8947 (4.0*10 ⁻⁴)	—	—	—	—	—	—	—	36155
—	—	—	25.1701 (1.8*10 ⁻³)	0.0406 (1.1*10 ⁻⁵)	1.1066 (2.0*10 ⁻⁴)	—	—	—	—	28842
Univariate mixed effects models										
36.4617 (3.9*10 ⁻³)	0.0285 (7.7*10 ⁻⁶)	1.4869 (2.6*10 ⁻⁵)	—	—	—	—	6.6263 (1.8*10 ⁻³)	—	—	35320
—	—	—	26.0732 (1.1*10 ⁻³)	0.0362 (4.6*10 ⁻⁶)	0.5841 (9.3*10 ⁻⁵)	—	—	3.9338 (1.0*10 ⁻³)	—	25831
Bivariate fixed effects model										
35.3462 (1.3*10 ⁻²)	0.0280 (2.2*10 ⁻⁵)	3.5968 (1.0*10 ⁻³)	25.3301 (4.0*10 ⁻³)	0.0396 (2.2*10 ⁻⁵)	1.1968 (6.1*10 ⁻⁴)	3.9187 (9.2*10 ⁻⁴)	—	—	—	62912
Bivariate mixed effects model										
35.9030 (9.8*10 ⁻³)	0.0293 (2.0*10 ⁻⁵)	2.3736 (2.2*10 ⁻³)	26.0039 (2.9*10 ⁻³)	0.0363 (1.2*10 ⁻⁵)	0.3615 (3.0*10 ⁻⁴)	0.6334 (6.6*10 ⁻⁴)	5.9592 (4.5*10 ⁻³)	3.6830 (2.6*10 ⁻³)	—	58815

Table 2. Estimates of parameters and AIC.

stand age and can be used to explain some aspects of stand evolution such as the mean values of diameter, height and stem volume and their coefficient of variation, diameter distribution, height distribution, dominant height and the mean stand basal area or stand volume (m³/ha).

Estimating results. The maximum likelihood estimator seeks (see Supplement Method) to make the conditional probability density function the most likely fit to the observed diameter and height estimation dataset $\{(d_1^i, h_1^i), (d_2^i, h_2^i), \dots, (d_n^i, h_n^i)\}$ at discrete times (ages) $\{t_1^i, t_2^i, \dots, t_n^i\}$ starting at the initial age $t_0 = 5$, diameter and height $X(t_0) = x_0 = (0.0, 1.3)^T$ from which a given set of SDEs (1) evolves. The parameter estimators, $\hat{\theta}^1 = \{\hat{\alpha}_d, \hat{\alpha}_h, \hat{\beta}_d, \hat{\beta}_h, \hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\sigma}_{22}\}$ and $\hat{\theta}^2 = \{\hat{\alpha}_d, \hat{\alpha}_h, \hat{\beta}_d, \hat{\beta}_h, \hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\sigma}_{22}, \hat{\sigma}_d, \hat{\sigma}_h\}$, were calculated by maximum likelihood and approximated maximum likelihood technique, respectively, realised by Eqs SM2, SM.8 and SM.9 (see Supplement Method), using the NLPsolve procedure in MAPLE 11³². The results of the parameter estimates, their standard deviations and the Akaike's Information Index (AIC)³³ are summarised in Table 2. The Akaike's Information Index is defined by:

$$AIC = -2 \cdot LL_s(\hat{\theta}^s), \quad s = 1, 2, \quad (12)$$

where $LL_s(\cdot)$ is defined in the Supplementary Method.

The AIC shows that including the random effects into the univariate and bivariate Vasicek-type SDE models improved goodness of fit of the univariate and bivariate conditional probability densities compared with the corresponding fixed effect parameters univariate and bivariate densities.

Random effects calibration. Calibration means that random effects are predicted using a supplementary dataset of observations taken from a new sampling unit. The theoretical diameter and height trends defined by Eqs 4 and 5 affect the conditional probability density function defined by Eq. 10. Hence, the random effects, ϕ_d^i, ϕ_h^i , for the i th stand, $i = 1, 2, \dots, K$, for the validation dataset can be calibrated for the diameter by:

$$\hat{\phi}_d^i = \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{d_j^i - \hat{\alpha}_d + (d_0 - \hat{\alpha}_d) \cdot \exp(-\hat{\beta}_d \cdot (t_j^i - t_0))}{1 - \exp(-\hat{\beta}_d \cdot (t_j^i - t_0))}, \quad (13)$$

where K is the number of plots in the validation dataset and m_i is the number of observed trees of the i th plot and for the height by:

$$\hat{\phi}_h^i = \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{h_j^i - \hat{\alpha}_h - (h_0 - \hat{\alpha}_h) \cdot \exp(-\hat{\beta}_h \cdot (t_j^i - t_0))}{1 - \exp(-\hat{\beta}_h \cdot (t_j^i - t_0))}. \quad (14)$$

Bivariate diameter and height distribution at a given age. The newly developed bivariate conditional probability density function for tree diameter and height is attractive for its simplicity and may be justified from an application perspective. To demonstrate that the data provided by the LNFI (2007–2010) of Scots pine trees do indeed follow the bivariate estimated probability density function defined by Eq. 10 we will use simple graphical techniques to illustrate the goodness of fit. Theoretical validation that a height and diameter dataset observed at discrete ages follows a bivariate probability density function is not easy and there is no simple statistical test. We graphically illustrate goodness of fit of the estimated bivariate density function (Eq. 10) with fixed effect and mixed effect parameters on three randomly selected plots from a validation dataset by plotting the estimated bivariate density functions, their contour plots and the observed datasets. Figure 1 shows the estimated bivariate mixed effect parameters estimated probability density functions for three randomly selected stands from the validation dataset as well as the estimated bivariate fixed effect parameters stationary (time t equates to

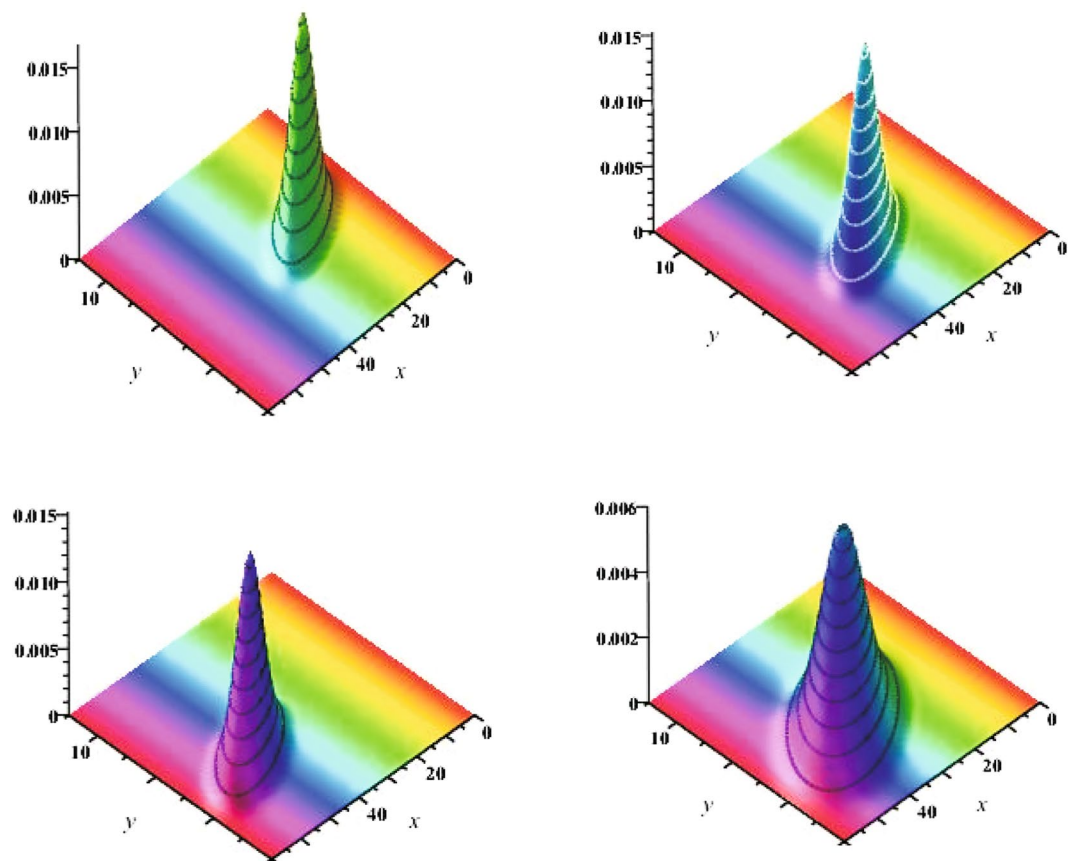


Figure 1. Bivariate estimated probability density functions defined by Eq. 10. First stand (mean of age – 44.0) – top left side; second stand (mean of age – 100.0) – top right side; third stand (mean of age – 145.0) – bottom left side; fixed effects stationary density – bottom right; x – diameter (cm); y – height (m).

infinity) probability density function (estimates of parameters $\hat{\theta}^1 = \{\hat{\alpha}_d, \hat{\alpha}_h, \hat{\beta}_d, \hat{\beta}_h, \hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\sigma}_{22}\}$ and $\hat{\theta}^2 = \{\hat{\alpha}_d, \hat{\alpha}_h, \hat{\beta}_d, \hat{\beta}_h, \hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\sigma}_{22}, \hat{\sigma}_d, \hat{\sigma}_h\}$ calculated by maximum likelihood procedure presented in the Supplementary Method are summarised in Table 2). The diameter and height random effects ϕ_d^i, ϕ_h^i ($i = 1, 2, 3$) for these three randomly selected stands were calibrated by Eqs 13 and 14, respectively.

The contour plots of the bivariate estimated mixed effects probability density function for three randomly selected stands from the validation dataset and the observed values, together with the contour plot of the bivariate estimated fixed effects stationary ($t = +\infty$) probability density function and the observed values from the same three randomly selected stands from the validation dataset are given in Fig. 2. Figure 2 shows that the mixed effect and fixed effect parameters bivariate estimated probability density functions well capture the main features of the data for three randomly selected stands from the validation dataset. The diameter and height random effects ϕ_d^i, ϕ_h^i for these three randomly selected stands were calibrated by Eqs 13 and 14, respectively.

The conditional distribution of the diameter at a given height and age and the conditional distribution of the height at a given diameter and age have univariate normal distributions $N(m_d^i(h, t), w_d^2(t))$ and $N(m_h^i(d, t), w_h^2(t))$, respectively. The mean, variance and coefficient of correlation can be calculated (using Eqs 4–10) in the following forms:

$$m_d^i(h, t) = E(D^i(t)|H^i(t) = h) = \mu_d^i(t) + \frac{\nu_{d,h}^2(t)}{\nu_h^2(t)}(h - \mu_h^i(t)), \quad (15)$$

$$w_d^2(t) = \text{Var}(D^i(t)|H^i(t) = h) = (1 - ro^2(t))\nu_d^2(t), \quad (16)$$

$$ro(t) = \frac{\nu_{d,h}^2(t)}{\sqrt{\nu_d^2(t)\nu_h^2(t)}}, \quad (17)$$

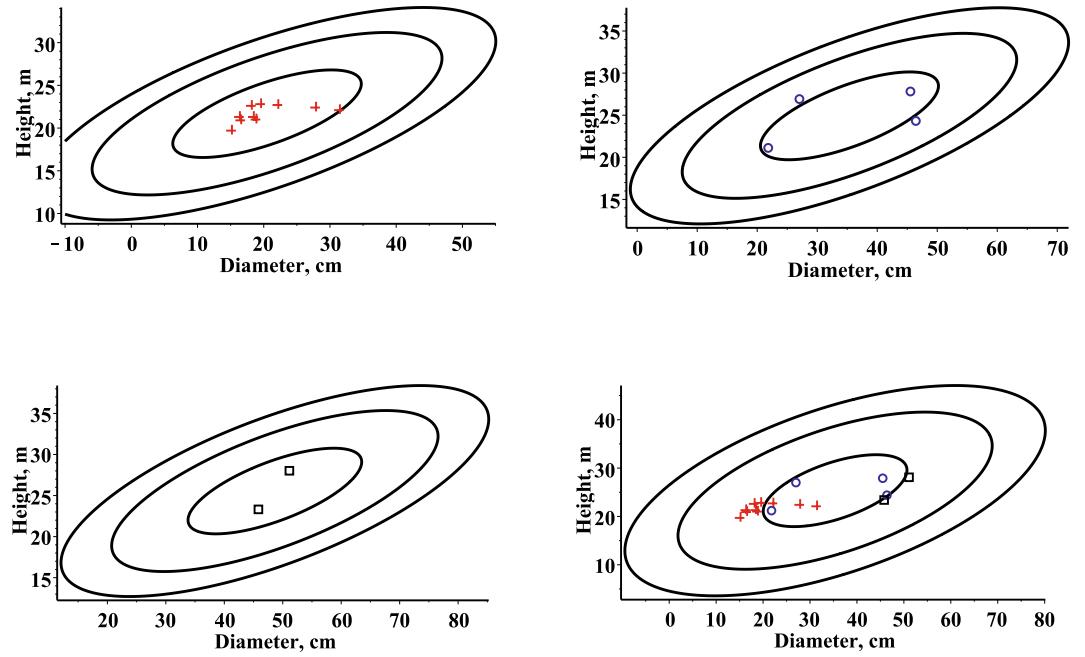


Figure 2. Contour plots (levels: 10^{-3} , 10^{-6} , 10^{-9}) of the estimated bivariate probability density function defined by Eq. 10. First stand (mean of age, diameter and height: 44.0, 20.5, 21.68) – top left side, observed values in red (cross); second stand (mean of age, diameter and height: 100.0, 35.4, 24.9) – top right side, observed values in blue (circle); third stand (mean of age, diameter and height: 145.0, 48.6, 25.6) - bottom left side, observed values in black (box); fixed effects stationary density – bottom right.

$$m_h^i(d, t) = E(H^i(t)|D^i(t) = d) = \mu_h^i(t) + \frac{\nu_{a,h}^2(t)}{\nu_d^2(t)}(d - \mu_d^i(t)), \tag{18}$$

$$w_h^2(t) = \text{Var}(H^i(t)|D^i(t) = d) = (1 - r\sigma^2(t))\nu_h^2(t). \tag{19}$$

In summary, Eqs 16 and 19 show us that the univariate conditional distribution of diameter (height) given height (diameter) has an age dependent variance, which is the same for each height (diameter). It is worth mentioning that the result on the height (diameter) independent variance of this study is influenced by properties of bivariate normal distribution. Figures 3 and 4 show the univariate conditional mixed effect parameters probability density functions at a given height and age (diameter and age) for three randomly selected stands from the validation dataset with the observed values, and the univariate conditional fixed effects parameters stationary ($t = +\infty$) probability density function at a given height and age (diameter and age).

Models of diameter and height. Diameter and height evolution can be formulated using a wide range of mathematical relationships from linearised fixed effect parameters regression equations to nonlinear mixed effect parameters generalised relationships. The mathematical technique for a system of uniform diameter and height regional functions is the approach known as the generalised model. The mixed effects regression models are able to achieve the same results as the generalised model³⁴.

Next, we summarise the results concerning the evolution of the diameter or height using univariate and bivariate fixed effects and mixed effects for the Vasicek-type growth models. Recall the models for the diameter and height growth:

univariate fixed effect parameters

$$d_1(t) = E(D(t)) = E(D(t)|D(t_0) = d_0) = \alpha_d + (d_0 - \alpha_d)e^{-\beta_d(t-t_0)}, \tag{20}$$

$$h_1(t) = E(H(t)) = E(H(t)|H(t_0) = h_0) = \alpha_h + (h_0 - \alpha_h)e^{-\beta_h(t-t_0)}, \tag{21}$$

univariate mixed effect parameters

$$d_2^i(t) = E(D^i(t)) = E(D(t)^i|D^i(t_0) = d_0) = \alpha_d + \phi_d^i - (d_0 - (\alpha_d + \phi_d^i))e^{-\beta_d(t-t_0)}, \tag{22}$$

$$h_2^i(t) = E(H^i(t)) = E(H(t)^i|H^i(t_0) = h_0) = \alpha_h + \phi_h^i + (h_0 - (\alpha_h + \phi_h^i))e^{-\beta_h(t-t_0)}, \tag{23}$$

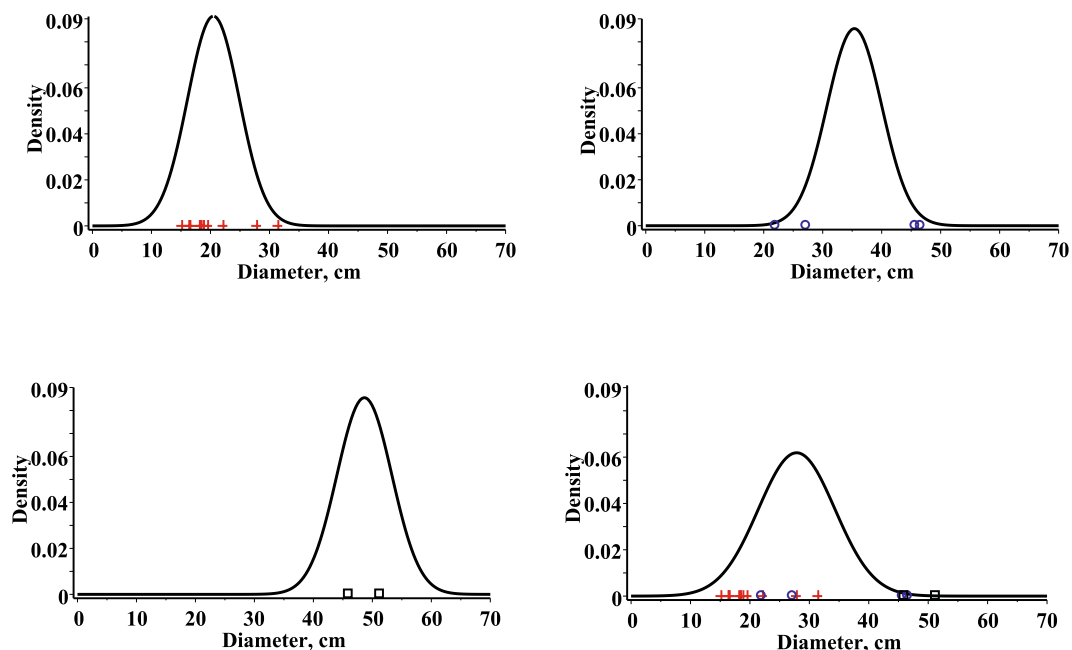


Figure 3. Univariate conditional probability density functions for diameters. First stand (mean of age and height: 44.0, 21.68) – top left side, observed values in red (cross); second stand (mean of age and height: 100.0, 24.9) – top right side, observed values in blue (circle); third stand (mean of age and height: 145.0, 25.6) - bottom left side, observed values in black (box); fixed effects stationary density – bottom right.

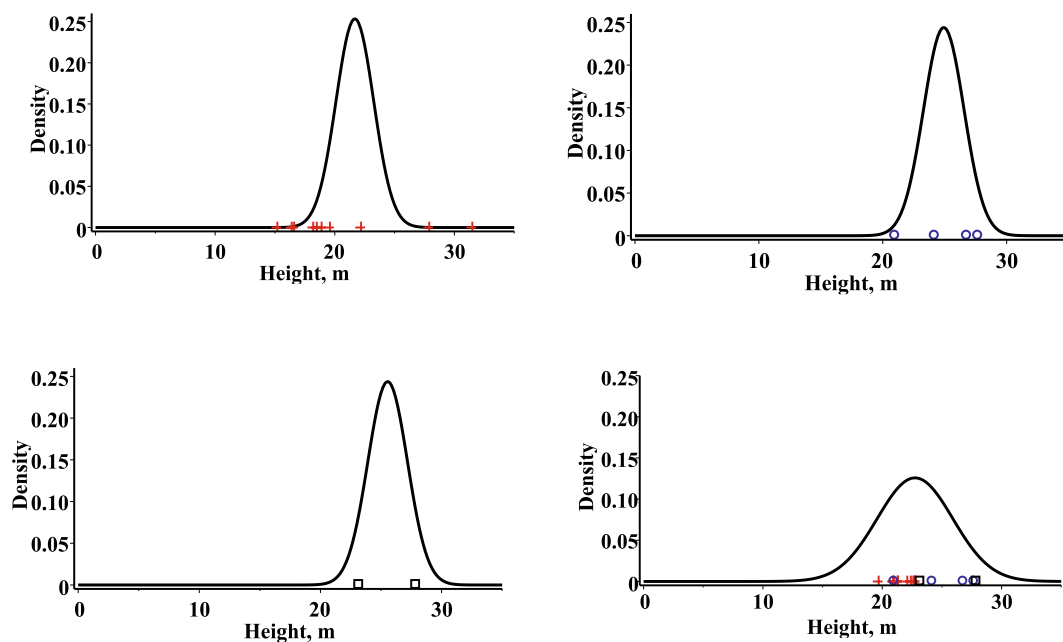


Figure 4. Univariate conditional probability density functions for heights. First stand (mean of age and diameter: 44.0, 20.5) – top left side, observed values in red (cross); second stand (mean of age and diameter: 100.0, 35.4) – top right side, observed values in blue (circle); third stand (mean of age and diameter: 145.0, 48.6) - bottom left side, observed values in black (box); fixed effects stationary density – bottom right.

bivariate fixed effect parameters

$$d_3(h, t) = E(D(t)|H(t) = h) = d_1(t) + \frac{v_{a,h}^2(t)}{v_h^2(t)}(h - h_1(t)), \tag{24}$$

Model	Estimation				Validation			
	B, m (PB, %)	AB, m (PAB, %)	AE, m	R ²	B, m (PB, %)	AB, m (PAB, %)	AE, m	R ²
Diameter								
1	0.0346 (-7.1912)	6.0611 (23.6162)	7.7593	0.1915	-0.1944 (-8.3678)	6.1214 (23.9642)	7.8347	0.2136
2	0.0147 (-4.5528)	4.2463 (16.4927)	5.4387	0.6027	-0.0025 (-3.7089)	4.0124 (15.5738)	5.2636	0.6447
3	0.0186 (-4.5137)	4.8389 (18.3712)	6.3144	0.4642	-0.1555 (-5.2887)	4.8765 (18.4629)	6.3604	0.4809
4	0.0212 (-2.3160)	3.0770 (11.6959)	3.9890	0.7863	-0.0009 (-1.8057)	3.0381 (11.5048)	4.0440	0.7903
Height								
1	0.0070 (-3.5060)	3.0014 (14.9750)	3.8506	0.2405	-0.0359 (-3.7091)	3.0078 (14.8898)	3.8385	0.2491
2	0.0018 (-1.0817)	1.4248 (6.9093)	1.8698	0.8209	-0.0004 (-0.8606)	1.4514 (6.8880)	1.9351	0.8091
3	- 0.0005 (-2.5673)	2.4610 (12.3000)	3.1442	0.4932	0.0177 (-2.4110)	2.4405 (12.0220)	3.1192	0.5034
4	-0.0060 (-0.6738)	1.0835 (5.2324)	1.3934	0.9005	6.2*10⁻⁶ (-0.5133)	1.1291 (5.3267)	1.4832	0.8876

Table 3. Statistical indexes for all models applied to the estimation and validation datasets. The best values of the statistical indexes are in bold, the mean prediction bias $\left(B = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)\right)$ and the percentage mean prediction bias $\left(PB = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i} \cdot 100\right)$, the absolute mean prediction bias $\left(AB = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|\right)$ and the percentage mean absolute prediction bias $\left(PAB = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\right)$, an adjusted root mean square error $\left(AE = \sqrt{B^2 + \frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{y}_i - B)^2}\right)$ and an adjusted coefficient of determination $\left(R^2 = 1 - \frac{n-1}{n-p} \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}\right)$. Here p is the number of parameters in the model (for fixed effects scenario p = 7 and for mixed effects scenario p = 9), n = $\sum_{i=1}^M n_i$ is the total number of observations used to fit the model, M is the number of stands, n_i is the number of measured trees in the *i*th stand, y_i , \hat{y}_i and \bar{y} are the measured, estimated and average values of the dependent variable (diameter or height).

$$h_3(d, t) = E(H(t)|D(t) = d) = h_1(t) + \frac{v_{d,h}^2(t)}{v_d^2(t)}(d - d_1(t)), \tag{25}$$

bivariate mixed effect parameters

$$d_4^i(h, t) = E(D^i(t)|H^i(t) = h) = d_2^i(t) + \frac{v_{d,h}^2(t)}{v_h^2(t)}(h - h_2^i(t)), \tag{26}$$

$$h_4^i(d, t) = E(H^i(t)|D^i(t) = d) = h_2^i(t) + \frac{v_{d,h}^2(t)}{v_d^2(t)}(d - d_2^i(t)). \tag{27}$$

Table 3 shows the predictive ability for all newly developed fixed- and mixed effect parameters diameter and height models for both estimation and validation datasets (using the estimates of parameters $\hat{\theta}^1 = \{\hat{\alpha}_d, \hat{\alpha}_h, \hat{\beta}_d, \hat{\beta}_h, \hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\sigma}_{22}\}$ and $\hat{\theta}^2 = \{\hat{\alpha}_d, \hat{\alpha}_h, \hat{\beta}_d, \hat{\beta}_h, \hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\sigma}_{22}, \hat{\sigma}_d, \hat{\sigma}_h\}$ presented in Table 2, and the random effects for diameter and height calibrated by Eqs 13 and 14, respectively). The estimated diameter evolution defined by Eq. 26 and the estimated height evolution defined by Eq. 27 were the best functions, with an increase in the coefficient of determination (R²) and decreases in the absolute prediction bias (AB) and the error (AE) for both estimation and validation datasets.

The estimated evolution of the diameter, $d = d_4^i(h, t)$, subject to the height and age, and the estimated evolution of the height, $h = h_4^i(d, t)$, subject to the diameter and age, for a randomly selected stand from the validation dataset are shown in Fig. 5, where we have assumed that the values of the predictor variable height (diameter) lie within a band around the mean $\mu_h^i(t)$ ($\mu_d^i(t)$) with a width of three standard deviations $\sqrt{v_h^2(t)}$ ($\sqrt{v_d^2(t)}$).

In Supplementary Fig. S1, the residuals of the diameter models defined by Eqs 20, 22, 24 and 26, and the LOWESS line (Locally Weighted Scatterplot Smoothing line) are plotted against the predicted diameter values. In Supplementary Fig. S2, the residuals of the height models defined by Eqs 21, 23, 25 and 27, and the LOWESS line are plotted against the predicted height values. Supplementary Figs S1 and S2 show that the residuals that were calculated using the bivariate mixed effects scenario are distributed more symmetrically around zero, with approximately constant variance, compared with the other scenarios. Therefore, the bivariate models incorporating the random effects of the stands were the best models for predicting diameter and height growth of individual Scots pine trees in the study area (see Supplementary Figs S1 and S2). A non-parametric smoothing line, called a

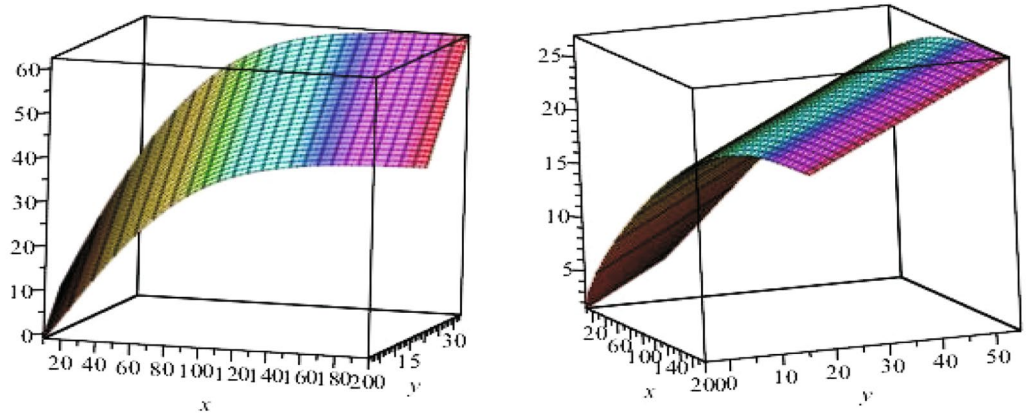


Figure 5. Estimated evolution of diameter and height for randomly selected stand (mean of age, diameter and height: 44.0, 20.5, 21.68). Left – diameter versus height and age (x – age, y – height); right – height versus diameter and age (x – age, y – diameter).

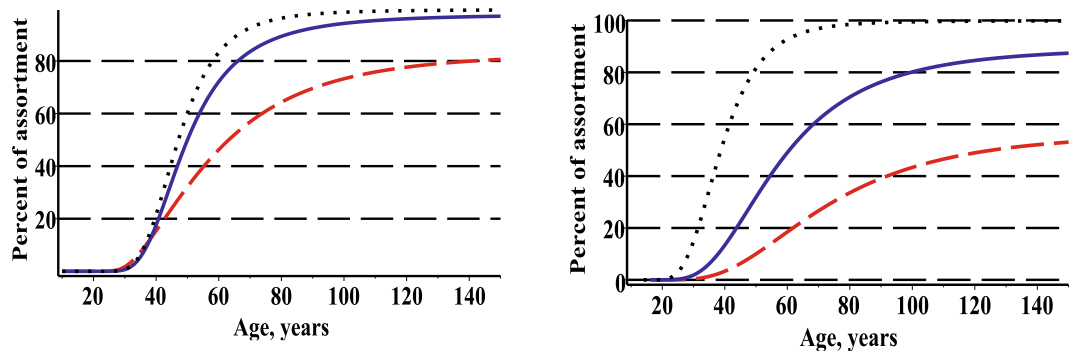


Figure 6. Estimated evolution of percentage of trees for three randomly selected stands and two scenarios. First stand – red on dash (mean of age, diameter and height: 44.0, 20.5, 21.68); second stand – blue on solid (mean of age, diameter and height: 100.0, 35.4, 24.9); third stand – black on dot (mean of age, diameter and height: 145.0, 48.6, 25.6); left scenario – diameter > 25, height > 20; right scenario – diameter > 30.

LOWESS line, shows a clear trend of predicted height; however, at the predicted diameter extremes the deviations are dictated by relatively little data.

The derived univariate conditional distribution of a diameter at a given height and age, and the derived univariate conditional distribution of a height at a given diameter and age, have normal distributions $N(m_d^i(h, t), w_d^2(t))$ and $N(m_h^i(d, t), w_h^2(t))$, respectively, and describe the knowledge of the predicted mean and uncertainty about the mean completely. These distributions can be used to describe the accuracy of predicted values by calculating the confidence intervals.

Stem structure. The bivariate conditional probability density function defined by Eq. 10 allows us to find the evolution of the percentage of trees in a particular stand that achieve fixed sizes for the diameter and height. Figure 6 shows the evolution of the percentage of two assortments for three randomly selected stands from the validation dataset.

Slenderness ratio. The slenderness coefficient is an important characteristic for indexing tree resistance to wind throw and snow damage. Figure 7 shows the evolution of the mean slenderness coefficient for three randomly selected plots. The slenderness of trees decreases with increasing stand age. For the bivariate fixed effect and mixed effect SDEs height and diameter distribution models the evolution of the slenderness is defined as follows:

$$SR(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{h}{d} \cdot f(d, h, t|\hat{\theta}) \cdot dd \cdot dh, \quad \hat{\theta} \in \left\{ \left(\hat{\theta}^1, 0, 0 \right), \left(\hat{\theta}^2, \hat{\phi}_d, \hat{\phi}_h \right) \right\}, \tag{28}$$

Mean and coefficient of variation of stem volume. For forestry applications, the most commonly used approach is to derive measures to estimate forest variables such as mean tree diameter, height and stem volume.

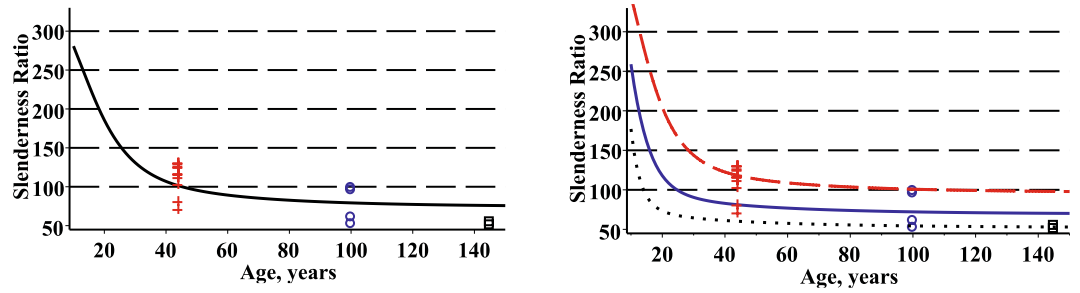


Figure 7. Estimated evolution of mean slenderness ratio for three randomly selected stands. Left – fixed effects scenario; right – mixed effects scenario; first stand – red on dash (mean of age, diameter and height: 44.0, 20.5, 21.68), observed values in cross); second stand – blue on solid (mean of age, diameter and height: 100.0, 35.4, 24.9), observed values in circle; third stand – black on dot (mean of age, diameter and height: 145.0, 48.6, 25.6), observed values in box.

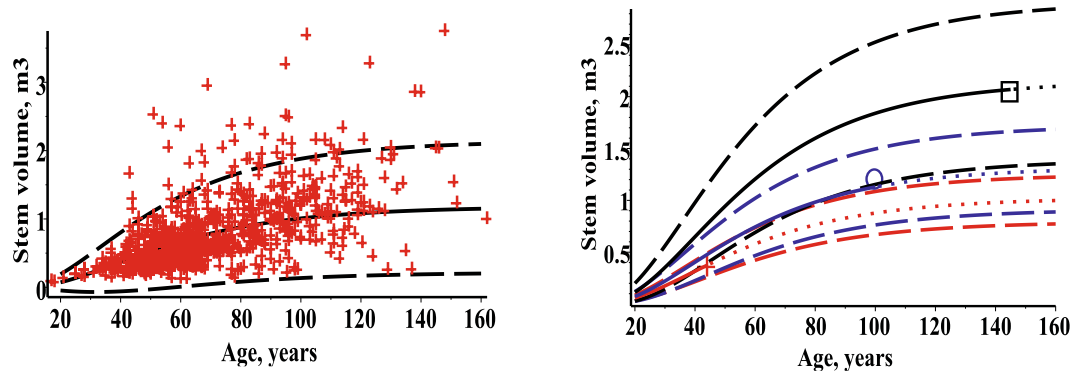


Figure 8. Estimated evolution of mean stem volume and mean $\pm 1.65 \cdot \text{standard deviation}$ (confidence bands). Left – fixed effect scenario: mean stem volume trend – solid; confidence bands – dash; observed mean stem volumes in cross. Right – mixed effect scenario for three randomly selected stands: first stand – red (mean of age, diameter and height: 44.0, 20.5, 21.68) and observed mean stem volume – cross; second stand – blue (mean of age, diameter and height: 100.0, 35.4, 24.9) and observed mean stem volume – circle; third stand – black (mean of age, diameter and height: 145.0, 48.6, 25.6) and observed mean stem volume – box. Mean stem volume trend from initial age to observed stand age – solid and from this age forward (forecast) – dot. Confidence bands – dash.

First, the bivariate fixed effects and mixed effects conditional probability density function of the diameter and height allows us to calculate the evolution of the mean stem volume in the following form:

$$\bar{V}(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} V(d, h) \cdot f\left(d, h, t \mid \hat{\theta}\right) \cdot dd \cdot dh, \quad \hat{\theta} \in \left\{ \left[\hat{\theta}^1, 0, 0 \right], \left[\hat{\theta}^2, \hat{\phi}_d, \hat{\phi}_h \right] \right\}, \tag{29}$$

where $V(d, h)$ is the individual stem volume regression function of power form³⁵:

$$V(d, h) = \beta_1 d^{\beta_2} h^{\beta_3}. \tag{30}$$

Estimates of the parameters β_1 , β_2 , and β_3 were calculated by a weighted least squares technique³⁶. The estimators and their standard deviations (in parenthesis) are, $\hat{\beta}_1 = 5.8 \cdot 10^{-5}$ ($5.8 \cdot 10^{-6}$), $\hat{\beta}_2 = 1.8801$ (0.028), $\hat{\beta}_3 = 0.9723$ (0.045). The estimated evolution of the mean stem volume in a stand for the fixed effect and mixed effect models and $\pm 1.65 \cdot SD_{\bar{V}}(t)$ confidence bands (if we assume normality then they covers 90%) are shown in Fig. 8 for all stands (fixed effect scenario) and three randomly selected stands (mixed effect scenario) from the validation dataset, respectively. Observed values of mean stem volumes in all stands were defined by the average of stem volumes calculated by the regression (Eq. 30). For the mixed effect scenario the mean stem volume estimate proves satisfactory, with the mean prediction bias (the percentage mean prediction bias) -0.0158 m^3 (-3.28%) and with the mean prediction absolute bias (the percentage mean prediction absolute bias) 0.0322 m^3 (5.07%). For the validation dataset the percent of the mean stem volume variation explained attains high levels too, 99.51%, for the mixed effect methodology.

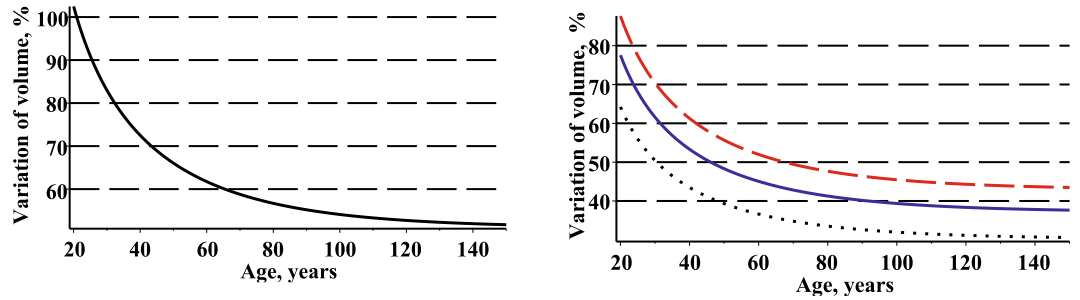


Figure 9. Coefficient of variation of stem volume for three randomly selected stands. Left – fixed effects scenario; right – mixed effects scenario; first stand – red on dash (mean of age, diameter and height: 44.0, 20.5, 21.68); second plot – blue on solid (mean of age, diameter and height: 100.0, 35.4, 24.9); third stand – black on dot (mean of age, diameter and height: 145.0, 48.6, 25.6).

Second, the bivariate fixed effects and mixed effects conditional probability density functions of the diameter and height allow us to estimate stem volume variability by the coefficient of variation. We estimate the evolution of the standard deviation of the stem volume in the following form:

$$SD_V(t) = \sqrt{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} V(d, h)^2 \cdot f(d, h, t | \hat{\theta}) \cdot dd \cdot dh - \bar{V}(t)^2}, \quad \hat{\theta} \in \left\{ \left(\hat{\theta}^1, 0, 0 \right), \left(\hat{\theta}^2, \hat{\phi}_d, \hat{\phi}_h \right) \right\}. \tag{31}$$

Hence, the coefficient of variation takes the following form:

$$CV(t) = \frac{SD_V(t)}{\bar{V}(t)} \cdot 100. \tag{32}$$

The coefficient of variation of stem volume defined by Eq. 32 is a statistical measure of the dispersion of the volumes calculated by Eq. 30 data points around the mean stem volume calculated by Eq. 29. The coefficient of variation of stem volume is a useful statistic for comparing the degree of variation of stem volume from one stand to another, even if the means are drastically different from one another. Figure 9 shows a plot of the coefficient of variation as a function of stand age using the mean trend and standard deviation functions of stem volume. The coefficient of variation of the stem volume evolves into a stationary coefficient of variation. The coefficient of variation based on stem volume decreases with an increase in stand age.

Stand volume per ha. Estimation of stand volume per ha is important for sustainable forest management. Methods to forecast stand volume per ha (or stem volume structure) for a short or long time period are necessary to ensure forest management schemes. Traditionally, estimating stand volume per ha includes several steps, such as the choice of stem volume regression model, the choice of a height-diameter model when tree height data are incomplete, the expansion of a sample volume to the number of trees it represents per unit area or the choice of stand volume per ha model developed by replacing the original tree variables into equivalent stand variables³⁷ (tree height are replaced by dominant height and diameter at breast height by quadratic mean diameter). The SDEs framework enables us to characterise a stand volume per ha as a function of any specified stand age, t , in the following form:

$$V_S(t) = m_N(t) \cdot \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} V(d, h) \cdot f(d, h, t | \hat{\theta}) \cdot dd \cdot dh, \quad \hat{\theta} \in \left\{ \left(\hat{\theta}^1, 0, 0 \right), \left(\hat{\theta}^2, \hat{\phi}_d, \hat{\phi}_h \right) \right\}, \tag{33}$$

here $m_N(t)$ is the number of trees per ha at a given age, t . The Vasicek-type univariate SDE that describes the development of the number of trees per ha, $N^i(t)$, is defined by $dN^i(t) = \beta(\alpha + \phi_N^i - N^i(t)) \cdot dt + \sigma \cdot W^i(t)$, $i = 1, 2, \dots, M$, $P(N^i(200) = \delta) = 1$, here δ is unknown parameter representing the number of trees at the age of the 200 years, ϕ_N^i is normally distributed random variable with zero mean and constant variance, $\phi_N^i \sim N(0; \sigma_N^2)$. The conditional distribution of the number of trees per ha at a given age has a univariate normal distribution $N(\mu_N^i(t); w_N^2(t))$, with mean $\mu_N^i(t) = \alpha + \phi_N^i - (\delta - (\alpha + \phi_N^i) \cdot e^{\beta(t-200)})$ and variance $w_N^2(t) = \frac{\sigma^2(1 - e^{2\beta(t-200)})}{2\beta}$. The estimates of the parameters and their standard deviations (in parenthesis) are: for the fixed effects scenario $\hat{\alpha} = 41215.9217$ (74.6637), $\hat{\beta} = 8.84 \cdot 10^{-5}$ ($1.62 \cdot 10^{-7}$), $\hat{\sigma} = 18.9385$ (0.0029), $\hat{\delta} = 10.0$ (0.1744); for the mixed effects scenario $\hat{\alpha} = 2547.4181$ (0.4148), $\hat{\beta} = 0.0012$ ($8.86 \cdot 10^{-8}$), $\hat{\sigma} = 5.8849$ (0.0015), $\hat{\delta} = 94.7886$ (0.0540), $\hat{\sigma}_N = 1516.9969$ (0.4449). Figure 10 shows the evolution of the stand volume per ha as a function of a stand age using the fixed effects scenario and the mixed effects scenario (for three randomly selected stands from the validation dataset).

Conclusions

We focus on discussion of the bivariate Vasicek-type stochastic process that drives the density evolution of the tree diameter and height structure via stand age. In this study, principal concepts of the bivariate mixed effect

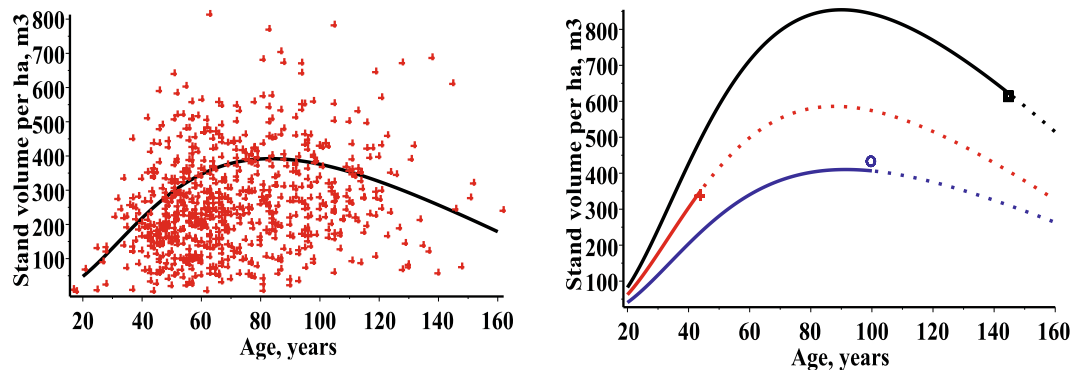


Figure 10. Estimated evolution of stand volume per ha (m^3) using fixed and mixed effect scenarios. Left – fixed effect scenario: stand volume trend – solid and observed stand volumes – crosses. Right – mixed effect scenario for three randomly selected stands: first stand – red (mean of age, diameter and height: 44.0, 20.5, 21.68) and observed stand volume – cross; second stand – blue (mean of age, diameter and height: 100.0, 35.4, 24.9) and observed volume – circle; third stand – black (mean of age, diameter and height: 145.0, 48.6, 25.6) and observed stand volume – box. Stand volume trend from initial age to observed stand age – solid and from observed age forward (forecast) – dot.

parameters Vasicek-type SDE in simple terms are presented and results of the diameter and height distribution, mean diameter, mean height, mean stem volume and stand volume per ha evolution via stand age are illustrated using the Lithuanian National Forest dataset from 2007 to 2010 for Scots pine trees. A new SDE framework for bivariate diameter and height distribution modelling is demonstrated using stand-level random effects. The newly developed conditional probability density functions are easily applied for the modelling of the evolution of the stand attributes via stand age.

The newly developed bivariate conditional probability density functions allowed us to calculate the evolution, in the forward and backward directions, of the mean diameter, height, dominant height, stem volume, stem volume of a stand and the uncertainties in these characteristics completely for a given stand age. The derived conditional probability density can be used to describe the accuracy of the predicted values (diameter, height and stem volume) in the sense of confidence bands.

It is easy to estimate fixed and random effects of the bivariate SDE using large datasets, which in this case was provided by the National Forest Inventories (one cycle). Moreover, the newly developed bivariate mixed effects conditional probability density function is an appropriate model for examination of diameter and height distributions of uneven-aged forest stands.

The present paper attempts to expand the scientific knowledge related to the evolution of tree diameter and height distributions in a forest stand by using SDEs, random effect technique and their possible applications to the modelling of the mean diameter, height, stem volume and stand volume per ha of a stand. Although this study has some limitations (e.g. estimation and validation datasets are not provided with remeasurement cycles of plots), these findings contribute to a better understanding of how stand growth and yield attributes are related to the stand age. Therefore, in our opinion, this study represents one of the first attempts to examine the evolution of growing stock resources, their structure and changes using the bivariate random process methodology and utilizing the data from one cycle of the LNFI. For a better understanding of the tree number and stand volume per ha dynamics remeasurement data from permanent plots (LNFI) could be utilized.

The trivariate (diameter, height and number of trees per ha) distributional model would be superior to the bivariate (diameter and height) distributional model in accordance to the underlying covariance structure driving changes in the tree diameter, height and number of trees per ha.

The number of trees per ha plays a central role for stand growth, but its representation in the LNFI dataset is limited by data availability. Missing data on initial density of stand (number of trees per ha at the age of $t_0 = 5$), restricted ability to develop a trivariate model.

Further research on the interplay between tree diameter, height and number of trees per ha using trivariate diffusion process and remeasurement data (2 or 3 cycles) from the LNFI would improve the consequences for the tree number and stand volume per ha evolution.

The analysis of the mixed effect SDE bivariate model by the specification of one random effect structure that used estimation based on a maximum likelihood procedure is not properly supported by the LNFI dataset. On the other hand, the model with the structure of all possible random effect components included may not converge.

References

1. Cajanus, W. Über die Entwicklung gleichartiger Waldbestände. *Acta For. Fenn.* **3**, 1–142 (1914).
2. Schreuder, H. T. & Hafley, W. L. A useful bivariate distribution for describing stand structure of tree heights and diameters. *Biometrics* **33**, 471–478 (1977).
3. Anfodillo, T. *et al.* An allometry-based approach for understanding forest structure, predicting tree-size distribution and assessing the degree of disturbance. *Proc. R. Soc. B* **280**, 20122375 (2012).
4. Meyer, W. H. Diameter distribution series in even aged forest stands. *Yale Univ. Sch. For. Bull.* **28**, 1–105 (1930).
5. Schnur, G. L. Diameter distributions for oldfield loblolly pine stands in Maryland. *J. Agric. Res.* **49**, 731–743 (1934).
6. Nelson, T. C. Diameter distribution and growth of loblolly pine. *For. Sci.* **10**, 105–115 (1964).

7. Bliss, C. L. & Reinker, K. A. A lognormal approach to diameter distributions in even-aged stands. *For. Sci.* **10**, 350–360 (1964).
8. Bailey, R. L. & Dell, T. R. Qualifying diameter distributions with the Weibull function. *For. Sci.* **19**, 97–104 (1973).
9. Wang, M. & Rennolls, K. Tree diameter modelling: introducing the logit-logistic distribution. *Can. J. For. Res.* **35**, 1305–1313 (2005).
10. Chen, C. M. & Rose, D. W. Direct and indirect estimation of height distributions in even-aged stand. *Minnesota Forestry Research Notes* **267**, 1–3 (1978).
11. Loetsch, F., Zöhrer, F. & Haller, K. *Forest Inventory, Vol. II* (BLV Verlagsgesellschaft, München, 1973).
12. Hafley, L. & Schreuder, H. T. Statistical distributions for fitting diameter and height data in even-aged stands. *Can. J. For. Res.* **7**, 481–487 (1977).
13. Mønness, E. The Power Normal distribution: application to forest stands. *Can. J. For. Res.* **41**, 707–714 (2011).
14. Wang, M. & Rennolls, K. Bivariate distribution modelling with tree diameter and height data. *For. Sci.* **53**, 16–24 (2007).
15. Mønness, E. The bivariate Power-Normal and the bivariate Johnson's system bounded distribution in forestry, including height curves. *Can. J. For. Res.* **45**, 307–313 (2015).
16. Plackett, R. L. A class of bivariate distributions. *J. Am. Statist. Assoc.* **60**, 516–522 (1965).
17. Li, F., Zhang, L. & Davis, C. J. Modeling the joint distribution of tree diameters and heights by bivariate generalized beta distribution. *For. Sci.* **48**, 47–58 (2002).
18. Gorgoso-Varela, J. J., García-Villabrille, J. D., Rojo-Alboreca, A., von Gadow, K. & Álvarez-González, J. G. Comparing Johnson's SBB, Weibull and Logit-Logistic bivariate distributions for modeling tree diameters and heights using copulas. *For. Syst.* **25**, eSC07 (2016).
19. Mandt, S., Sadri, S., Houck, A. A. & Türeci, H. E. Stochastic differential equations for quantum dynamics of spinboson networks. *New J. Phys.* **17**, 053018 (2015).
20. Mollgaard, A. & Mathiesen, J. Emergent user behavior on Twitter modelled by a stochastic differential equation. *PLoS ONE* **10**(5), e0123876 (2015).
21. Mwaffo, V., Butail, S. & Porfiri, M. *In-silico* experiments of zebrafish behaviour: modeling swimming in three dimensions. *Sci. Rep.* **7**, 39877 (2017).
22. Rupšys, P. & Petrauskas, E. Analysis of height curves by stochastic differential equations. *Int. J. Biomath.* **5**(5), 1250045 (2012).
23. Rupšys, P. & Petrauskas, E. The bivariate Gompertz diffusion model for tree diameter and height distribution. *For. Sci.* **56**, 271–280 (2010).
24. Rupšys, P. Generalized fixed-effects and mixed-effects parameters height-diameter models with diffusion processes. *Int. J. Biomath.* **8**(5), 1550060 (2015).
25. Rupšys, P. Stochastic mixed-effects parameters Bertalanffy process, with applications to tree crown width modeling. *Math. Prob. Eng.* **2015**, 375270 (2015).
26. Rupšys, P. Height-diameter models with stochastic differential equations and mixed-effects parameters. *J. For. Res.* **20**, 9–17 (2015).
27. Rupšys, P. New insights into tree height distribution based on mixed effects univariate diffusion processes. *PLoS ONE* **11**(12), e0168507 (2016).
28. Rupšys, P. The use of copulas to practical estimation of multivariate stochastic differential equation mixed effects models. *AIP Conference Proceedings* **1684**, 080011 (2015).
29. Uhlenbeck, G. E. & Ornstein, L. S. On the theory of Brownian motion. *Phys. Rev.* **36**, 823–841 (1930).
30. Itô, K. On stochastic processes. *Jap. J. Math.* **18**, 261–301 (1942).
31. García, O. Cohort aggregation modelling for complex forest stands: Spruce–aspen mixtures in British Columbia. *Ecol. Model.* **343**, 109–122 (2017).
32. Monagan, M. B. *et al. Maple Advanced Programming Guide* (Maplesoft, Printed in Canada, 2007).
33. Akaike, H. A new look at the statistical model identification. *IEEE T. Automat. Contr.* **19**, 716–723 (1974).
34. Adamec, Z. Comparison of linear mixed effects model and generalized model of the tree height-diameter relationship. *J. Forest Sci.* **61**, 439–447 (2015).
35. Schumacher, F. X. & Hall, F. D. S. Logarithmic expression of timber tree volume. *J. Agric. Res.* **47**, 719–734 (1933).
36. Petrauskas, E., Bartkevičius, E., Rupšys, P. & Memgaudas, R. The use of stochastic differential equations to describe stem taper and volume. *Baltic For.* **19**, 43–151 (2013).
37. Tang, X. *et al.* Development of a compatible taper function and stand-level merchantable volume model for Chinese fir plantations. *PLoS ONE* **11**(1), e0147610 (2016).

Acknowledgements

We are thankful to the Lithuanian Association of Impartial Timber Scalpers for the financial support.

Author Contributions

E.P. organized and interpreted the data. P.R. worked on the theoretical analysis and performed statistical analysis. P.R. and E.P. wrote and edited the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-16100-2>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017