Re-examination of the Taper Models by Stochastic Differential Equations

PETRAS RUPŠYS1,3, EDMUNDAS PETRAUSKAS1, EDMUNDAS BARTKEVIČIUS2, ROMAS MEMGAUDAS1

1Department of Forest Management, 2Department of Ecology, 3Department of Mathematics and Informatics
Lithuanian University of Agriculture
Studenų g. 11, Akademija, Kaunas district. LT – 53361
LITHUANIA
petras.rupsys@lzua.lt

Abstract: This paper introduces the application of stochastic differential equations for the tree taper modeling. We propose to model taper curves using stochastic differential equations that are deduced from the standard deterministic growth function by adding random variations to the growth dynamics. It is assumed that the conditional distribution of the taper diffusion process has an explicit form. We illustrate the efficiency of our method using the logistic and Vasicek diffusion processes to model data on Scots pine trees.

Key-words: Stochastic differential equations; taper curves; fixed and mixed models; log-likelihood functions.

1 Introduction
Taper equations are widely used in forestry to estimate diameter at any given height along a tree bole and therefore to calculate total or merchantable stem volume. One crucial element in these models is the functional response describing the relative diameter of tree stem consumed per relative height for given quantities of diameter at breast height \( D \) and total tree height \( H \). The most studied of the taper relations is from simple taper functions to more complex forms [1], [2]. Taper curve data consist of repeated measurements of a continuous diameter growth process over height of individual trees. These longitudinal data have two characteristics that complicate their statistical analysis: a) within-individual tree correlation that appears with data measured on the same tree and b) independence but extremely high variability between the experimental taper curves of the different trees. Mixed models provide one of powerful tools to analysis of longitudinal data. These models incorporate the variability between individual trees by means of the expression of the model's parameters and in terms of both fixed and random effects. Each parameter in the model may be represented by a fixed effect that stands for the mean value of the parameter as well as a random effect that expresses the difference between the value of the parameter fitted for each specific tree and the mean value of the parameter - the fixed effect. Random effects are conceptually random variables. They are modeled as such in terms of describing their distribution. This helps to avoid the problem of overparameterisation. A large number of mixed-effect taper models have been completed, and the study is still one of the important issues in progress [3], [4], and references therein.

The increasing popularity of mixed-effects lies in their ability to model total variation, splitting it into its within- and between-individual tree components. We propose to model these variations using stochastic differential equations that are deduced from the standard deterministic growth function by adding random variations to the growth dynamics [5]-[9]. We thus consider stochastic differential equation mixed-effect models whose drift and diffusion terms can depend linearly or nonlinearly on state variables and random effects following the normal distribution.

The aim of this study is to put forward the advantages of using stochastic differential equations in the analysis of taper models and to
show how an adequate model can be made. In this paper attention is restricted to homogeneous stochastic differential equations in the logistic, and Vasicek type [8], whose solution is the regression term of the mixed model.

2 Stochastic Differential Equation Model

Consider a one-dimensional continuous process \( Y(x) \) evolving in \( M \) different experimental units (trees) randomly chosen from a theoretical population (tree species). We suppose that dynamic of relative diameter \( x = \frac{h}{H} \) is expressed by a stochastic differential equation, where \( d \) is the diameter outside bark at any given height \( h \), \( D \) is the diameter at breast height outside bark, \( H \) is the total tree height from ground to tip. Stochastic differential equations approach adds dynamic noise, allows intra-individual variations through fluctuation of transfer rates over relative height, and accounts for not-well understood complex taper process. The first model of relative diameter dynamic is defined in the following logistic form [8],

\[
\begin{align*}
  & dy^i(x) = \left( \alpha + b^i \right) y^i(x) - \beta y^i(x) \right] dx + \sigma dW^i(x) \\
  & P(Y^i(x_0) = y_0^i, 1, i = 1, \ldots, M,)
\end{align*}
\]

(1)

where \( Y^i(x) \) is the value of the process at relative height \( x \geq x_0^i \), \( \alpha, \beta, \sigma \) are fixed effects parameters (the same for the entire population of trees), and \( b^i \) is a random effects parameter (treespecific). The normal density function of the parameter \( b^i \) is denoted by \( p_b(y_0^i | \sigma_b) \). The \( W^i(x) \), \( i = 1, \ldots, M \) are standard Brownian motions. The \( W^i(x) \) and \( b^i \) are assumed mutually independent for all \( 1 \leq i, j \leq M \). The second model of relative diameter dynamic is defined in the following Vasicek form [8],

\[
\begin{align*}
  & dy^i(x) = \beta \left( \alpha + b^i \right) - \gamma y^i(x) \right] dx + \sigma dW^i(x) \\
  & P(Y^i(x_0) = y_0^i, 1, i = 1, \ldots, M,)
\end{align*}
\]

(2)

Assume that tree \( i \) is observed at \( n_i + 1 \) discrete relative height points \( (x_{0i}, x_{1i}, \ldots, x_{ni}) \) \( i = 1, \ldots, M \). Let \( y_i^j \) be the vector of responses (relative diameter) for tree \( i \), \( y_i^j = (y_{0i}, y_{1i}, \ldots, y_{ni}) \), where \( y^i(x_j) = y_j^i \), \( y = \left( y_1^1, y_1^2, \ldots, y_M^M \right) \) be the \( n \)-dimensional total relative diameter vector, \( n = \sum_{i=1}^{M} (n_i + 1) \), and let \( \Delta_j = x_j^i - x_{j-1}^i \) for the relative height distance between the observations \( x_j^i \) and \( x_{j-1}^i \). Therefore, we need to estimate \( \alpha, \beta, \sigma, \sigma_b \) using simultaneously all the data in \( y \). As we see, the specific values of the random effects parameters \( b^i \) are not of interest, but only the estimation of the parameter \( \sigma_b \) characterizing their distribution.

The models proposed in this paper use one treespecific prior relative diameter \( y_0^i \) (this known initial condition additional needs upper stem diameter measured at a stem height of 0 m). The transition probability density function of relative diameter stochastic processes \( Y^i(x_j^i), x_j^i \in [0;1], \)

\( i = 1, \ldots, M, j = 0, \ldots, n_i \) defined by Eqs. (1)-(2), can be deduced in the following form: for the logistic model

\[
\begin{align*}
  & p_y(y_j^i | x_j^i, y_0^i, \alpha, \beta, \sigma, b^i) \\
  & = \frac{2 (\alpha + b^i)(y_j^i)}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \frac{y_j^i e^{\left( \alpha + b^i \right) y_j^i}}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \frac{y_j^i + 1}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \frac{y_j^i + 1}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \frac{y_j^i + 1}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \frac{y_j^i + 1}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \\
  & \times \exp \left[ \frac{-2 (\alpha + b^i)(y_j^i)}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \frac{y_j^i e^{\left( \alpha + b^i \right) y_j^i}}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \frac{y_j^i + 1}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \frac{y_j^i + 1}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \frac{y_j^i + 1}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \frac{y_j^i + 1}{\sigma^2 \left( 1 - e^{-(-b^i)} \right)} \right].
\end{align*}
\]

(3)

\[
\begin{align*}
  & p_y(y_j^i | x_j^i, y_0^i, \alpha, \beta, \sigma, b^i) \\
  & = \frac{1}{s^2} \exp \left[ \frac{-\left( y_j^i - \left( \alpha + b^i \right) \right)^2}{s^2} \right] \\
  & s^2 = \frac{\sigma^2}{\beta \left( 1 - e^{-b^i} \right)}.
\end{align*}
\]

(4)
The mean and variance functions \( m(x') \), \( v(x') \)
\((x' \text{ is the relative height of } i\text{'th tree})\) of the stochastic processes (1)-(2) are defined by

\[
m_i(x') = \frac{2(\alpha + b^i)}{\gamma \sigma^2 [1 - e^{-(\alpha + b^i) x'}]} \Phi \left( 1, y_0 + 1, -\frac{2(\alpha + b^i)}{y_0 \sigma^2 [1 - e^{-(\alpha + b^i) x'}]} \right) \quad (5)
\]

\[
v_i(x') = \frac{1}{\gamma} \left( \frac{2(\alpha + b^i)}{\sigma^2 [1 - e^{-(\alpha + b^i) x'}]} \right)^2 \times
\frac{1}{y_0 - 1} \Phi \left( 2, y_0 + 1, -\frac{2(\alpha + b^i)}{y_0 \sigma^2 [1 - e^{-(\alpha + b^i) x'}]} \right)
\times \Phi \left( 2, y_0 + 1, -\frac{2(\alpha + b^i)}{y_0 \sigma^2 [1 - e^{-(\alpha + b^i) x'}]} \right)
\quad (6)
\]

\[
\gamma = \frac{2 \beta}{\sigma^2} + 1
\]

for the logistic model ( \( \Phi() \) is a Kummer function); and

\[
m_\nu(x') = (\alpha + b^i) + (y_0^i - (\alpha + b^i)) e^{-\beta x'}, \quad (7)
\]

\[
v_\nu(x') = \frac{\sigma^2}{2 \beta} \left( 1 - e^{-2 \beta x'} \right)
\]

for the Vasicek model.

We consider the models (1)-(2) in two aspects. First, the maximum likelihood function is derived for fixed effects models (in this case the parameter of random effect \( b' \) is assumed to be equal its expected value \( \mathbb{E}(b') = 0, i = 1, ..., M \)). Second, the maximum likelihood function is derived for mixed models.

The log-likelihood function for fixed effects models can be defined as

\[
L_i(x') = \prod_{i=1}^{M} \left[ \prod_{j=1}^{n_i} \ln \left( p(y_j', x_j') \mid y_0^i, \alpha, \beta, \sigma, 0 \right) \right], \quad (8)
\]

where density function \( p(y_j', x_j') \mid y_0^i, \alpha, \beta, \sigma, 0 \) takes a type-form from (3) or (4).

The maximum likelihood function for mixed models takes the following form [10]

\[
L_i(x') = \prod_{i=1}^{M} \left[ \prod_{j=1}^{n_i} \ln \left( p(y_j', x_j') \mid y_0^i, \alpha, \beta, \sigma, b' \right) \right] \quad (9)
\]

Here, \( p_b(b' \mid \sigma_B) \) is the normal density of the random effect. Unfortunately, the integral in (9) has not a closed form solution. Since analytic expression for the integrand in (9) is known the Laplace method may be used [10]. Let, define function

\[
f(b' \mid \alpha, \beta, \sigma, \sigma_B)
\]

\[
= \sum_{j=1}^{n} \ln \left( p(y_j', x_j') \mid y_0^i, \alpha, \beta, \sigma, b' \right) + \ln(p_b(b' \mid \sigma_B)) \quad (10)
\]

The integral \( \int e^{(\theta(b)) \mid \alpha, \beta, \sigma, \sigma_B} \cdot db' \), by a second-order

Taylor series expansion can be approximated as the Laplace approximation [11]

\[
\ln \int e^{(\theta(b)) \mid \alpha, \beta, \sigma, \sigma_B} \cdot db' \approx f(b' \mid \alpha, \beta, \sigma, \sigma_B) + \ln(2\pi) - \frac{1}{2} \ln \left( H(b' \mid \alpha, \beta, \sigma, \sigma_B) \right) \quad (11)
\]

where

\[
\hat{b}' = \arg \max_{b'} f(b' \mid \alpha, \beta, \sigma, \sigma_B), \quad (12)
\]

\[
H(b' \mid \alpha, \beta, \sigma, \sigma_B) = \partial^2 f(b' \mid \alpha, \beta, \sigma, \sigma_B) \bigg/ \partial b' \partial b' \bigg|_{b' = \hat{b}'} \quad (13)
\]

The log-likelihood function is approximately given by

\[
L_i(x') \approx \prod_{i=1}^{M} \left[ \prod_{j=1}^{n_i} \ln \left( f(b' \mid \alpha, \beta, \sigma, \sigma_B) + \ln(2\pi) \right) \right] \quad (14)
\]

The maximization of \( L_i(x') \) is a nested optimization problem. The internal optimization step estimates the \( b' \) for every tree. The external
optimization step maximizes \( LL(\alpha, \beta, \sigma, \sigma_Q) \) after plugging the \( \hat{b} \) into (14).

In this paper segmented stochastic taper process which consists of the logistic and Vasicek stochastic growth models takes the following form

\[
dY(x) = \left[ (\alpha + b)Y(x) - \beta Y(x)^2 \right] dx + \sigma Y(x) dW(x), \quad x \leq 0.52, \quad (15)
\]

\[
\frac{\alpha(\alpha + b) - Y(x)}{\beta} dx + \alpha dW(x), \quad x > 0.52
\]

Our proposed taper model consists of the logistic and Vasicek diffusion processes. For the segmented taper diffusion model defined by equation (15) a joint point was used at 0.52, as the fit statistics obtained the best values.

3 Results and Discussion

We focus on the modeling of Scots pine tree data set. Data are noisy diameter measurements of \( n = 1923 \) Scots pine trees at heights \( h = 0, 1, 1.3, 3, 5, \ldots \) meters. The corresponding data set are presented on Figure 1. Such a data set has been previously analyzed [12], who concluded that, among the standard regression taper models, the q-exponential segmented model is the most appropriate one.

![Fig. 1. Observed relative diameters via relative heights.](image)

For the fixed effect taper diffusion model all the parameters of stochastic differential equations (1), (2) were estimated simultaneously. A MAPLE program was used to carry out calculations. Figure 2 shows the residuals plotted against predictions of the diameter outside bark at any given height. Graphical diagnostics of the residuals for the diameter predictions showed that the residuals of the segmented diffusion taper model (15), defined by stochastic differential equations (1) and (2), had more homogeneous variance than other commonly used taper models [12]. Taper profiles for three randomly selected Scots pine trees with diameters outside the bark at breast height of 12.5, 24.5 and 38.2 cm, and total tree heights of 7.0, 24.0 and 33.2 m, respectively, are plotted in Figure 3. It is clear that all tree profiles followed the stem data very closely.

![Fig. 2. Residuals for the segmented diffusion taper model.](image)

4 Conclusion

A new taper model was developed using stochastic differential equations. An estimation method for taper model defined by stochastic differential equations, incorporating random effects, has been adjusted and evaluated through simulations.

Comparing fixed effects and mixed effects approaches, it was concluded that the mixed effect model fit did not differ conspicuously from the fixed effect model in estimating diameter outside bark at any given height.

Acknowledgement

This work was supported in part by the Lithuanian Association of Impartial Timber Scalers.

References:

[4] Y. Yang, S. Huang, S.X. Meng, Development of a Tree-specific Stem Profile Model for


