Evaluation of Stochastic Differential Equations Approach for Predicting Individual Tree Taper and Volume

Edmundas Bartkevičius, Edmundas Petrauskas, Petras Rupšys, and Graciano Russetti

Abstract—An approach combining information generating from different stochastic differential equations are recognized for improving predictive quality of stem profile (taper). The stochastic differential equations stem taper models were fitted to a data set of Scots pine trees collected across the entire Lithuanian territory. Comparison of the predicted stem taper and stem volume with those obtained using regression based models showed a predictive power to the stochastic differential equations models.

Index Terms—Stochastic differential equation, Gompertz, stem taper, height, diameter outside bark, volume

I. INTRODUCTION

PREDICTIVE forestry is a specific application of the field of mathematical modelling for describing the behavior of an individual tree and stand under a given set of environmental conditions. Traditionally, the relationship between volume, height and diameter has been modelled based on simple linear and nonlinear regression. The base assumption of regression models that observed variations from regression curve are constant at different values of a covariate would be realistic if the variations were due to measurement errors. Instead, it is unrealistic, as the variations are due to random changes on growth rates induced by environmental random perturbations. Stochastic differential equation models do not have such weakness. There is a long history characterizing the stem profile (taper) of trees. Mathematically defining stem taper is necessary for the accurate prediction of stem volume. Taper equations do just this and are important to foresters and forest scientists because they provide a flexible alternative to conventional volume equations. These 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 stem taper relations is from simple taper functions to more complex forms [1]-[7]. 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 modelled 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 [5]-[7].

The increasing popularity of mixed-effects models 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 [8]-[15]. We thus consider stochastic differential equation models whose drift and diffusion terms can depend linearly or nonlinearly on state variables. Although numerous sophisticated models exist for stem profile, relatively few models have been produced using stochastic differential equations [16].

Our main contribution is to expand stem taper and stem volume models by using stochastic differential equations and to show how an adequate model can be made. In this paper attention is restricted to homogeneous stochastic differential equations in the Gompertz, Geometric Brownian Motion and Ornstein-Uhlenbeck type.
II. STOCHASTIC DIFFERENTIAL EQUATIONS MODELS

Taper Models

Consider a one-dimensional stochastic process \( Y(x) \) evolving in \( M \) different experimental units (e.g. trees) randomly chosen from a theoretical population (tree species). We suppose that dynamics of relative diameter \( Y^i = \frac{d}{dx} Y^i \) via relative height \( x^i = \frac{h^i}{H^i} \) \( (x^i \in [0,1]) \) is expressed by a SDE, where \( d \) is the diameter outside bark at any given height \( h \), \( D^i \) is the diameter at breast height outside bark of \( i^{th} \) tree, \( H^i \) is the total tree height from ground to tip of \( i^{th} \) tree. In this paper is used a class of the Ito SDEs that are reducible to an Ornstein-Uhlenbeck process. The first model of relative diameter dynamic is defined in the following Gompertz form \([10], [12]\)

\[
dY^i(x^i) = \left[ \alpha_G Y^i(x^i) - \beta_G Y^i(x^i) \ln(Y^i(x^i)) \right] dx^i + \sigma_G Y^i(x^i) dW^i(x^i)
\]

\[
P(Y^i(x^i_0) = y^i_0) = 1, \quad i = 1, \ldots, M
\]

(1)

where \( Y^i(x^i) \) is the value of the diameter growth process at relative height \( x^i \geq x^i_0 \); \( \alpha_G, \beta_G, \sigma_G \) are fixed effects parameters (the same for the entire population of trees); \( y^i_0 \) is non-random initial relative diameter. The \( W^i(x^i) \), \( i = 1, \ldots, M \) are mutually independent standard Brownian motions. Intuitively, we interpret the terms \( W^i(x^i) \), \( i = 1, \ldots, M \) as ecological and environmental noises. The second model of relative diameter dynamic is defined in the following Geometric Brownian Motion form

\[
dY^i(x^i) = \alpha_{GB} Y^i(x^i) dx^i + \sigma_{GB} Y^i(x^i) dW^i(x^i)
\]

\[
P(Y^i(x^i_0) = y^i_0) = 1, \quad i = 1, \ldots, M
\]

(2)

where \( \alpha_{GB}, \sigma_{GB} \) are fixed effects parameters (the same for the entire population of trees). The third model of relative diameter dynamic is defined in the following Ornstein-Uhlenbeck form

\[
dY^i(x^i) = \left( \alpha_{OU} - \frac{Y^i(x^i)}{\beta_{OU}} \right) dx^i + \sigma_{OU} Y^i(x^i) dW^i(x^i)
\]

\[
P(Y^i(x^i_0) = y^i_0) = 1, \quad i = 1, \ldots, M
\]

(3)

where \( \alpha_{OU}, \beta_{OU}, \sigma_{OU} \) are fixed effects parameters (the same for the entire population of trees).

In this paper is used a segmented stochastic taper process which consists of three different SDEs defined by (1)-(3). It conforms to the paradigm of stem taper curve that marks three different stem sections along thebole (two points of inflection): the lower section corresponding to a neiloid shape, the middle section corresponding to a parabolic shape, and the upper section corresponding to a conic shape. Max and Burkhart \([17]\) proposed a segmented polynomial model that uses two joining points to link three different stem sections. Following this idea the stem taper SDEs models (with two joining points: 0.15, 0.75) are defined in the following two different forms (Gompertz-GBM-OU)

\[
dY^i(x^i)
\]

\[
\left\{ \begin{align*}
&\left[ \alpha_G Y^i(x^i) - \beta_G Y^i(x^i) \ln(Y^i(x^i)) \right] dx^i \\
&+ \sigma_G Y^i(x^i) dW^i_1(x), \quad x^i \leq 0.15
\end{align*} \right.
\]

\[
\alpha_{GB} Y^i(x^i) dx^i \\
+ \sigma_{GB} Y^i(x^i) dW^i_2(x), \quad 0.15 < x^i \leq 0.75
\]

(4)

where \( P(Y^i(x^i_0) = y^i_0) = 1, \quad i = 1, \ldots, M \) (stem butt is fixed),

\[
dY^i(x^i)
\]

\[
\left\{ \begin{align*}
&\left[ \alpha_G Y^i(x^i) - \beta_G Y^i(x^i) \ln(Y^i(x^i)) \right] dx^i \\
&+ \sigma_G Y^i(x^i) dW^i_1(x), \quad x^i \leq 0.15
\end{align*} \right.
\]

\[
\alpha_{OU} - \frac{Y^i(x^i)}{\beta_{OU}} dx^i \\
+ \sigma_{OU} Y^i(x^i) dW^i_3(x), \quad x^i > 0.75
\]

(5)

where \( P(Y^i(x^i_0) = y^i_0) = 1, \quad i = 1, \ldots, M \) (stem butt and top are fixed). Assume that tree \( i \) is observed at \( n_i + 1 \) discrete relative height points \( (x^i_0,x^i_1,\ldots,x^i_n) \) \( i = 1, \ldots, M \). Let \( y^i \) be the vector of responses (relative diameter) for tree \( i \), \( y^i = (y^i_0,y^i_1,\ldots,y^i_n) \), where \( y^i(x^i_j) = \ln(y^i_j) \), \( y = (y^1,y^2,\ldots,y^M) \) be the \( n \)-dimensional total relative diameter vector, \( n = \sum_{i=1}^{M}(n_i + 1) \). Therefore, we need to estimate fixed-effects using simultaneously all the data in \( y \). Both models proposed in this paper use one tree-specific prior relative diameter \( y^i_0 \) (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 \( Y^i(x^i_j) \), \( x^i_j \in [0,1], \quad i = 1, \ldots, M, \quad j = 0, \ldots, n_i \) defined by Eqs. (1)-(3), can be deduced in the following form: for the Gompertz stem taper model

\[
p_G \left( y^i_0, y^i_j; \alpha_G, \beta_G, \sigma_G \right) = \frac{1}{\sqrt{2\pi\alpha_G(x^i_j)} \beta_G \sigma_G} \exp \left\{ -\frac{1}{2\beta_G} \left[ \ln(y^i_j) - \mu_G(x^i_j|\alpha_G, \beta_G, \sigma_G) \right]^2 \right\}
\]
where
\[ 
\mu_G(x^i_j|\alpha_G, \beta_G, \sigma_G) = y_0^i e^{-\beta_G x^i_j} + \frac{1 - e^{-\beta_G x^i_j}}{\beta_G} \left( \alpha_G - \frac{\sigma_G^2}{2} \right) 
\]
\[ 
\nu_G(x^i_j|\beta_G, \sigma_G) = \frac{1 - e^{-2\beta_G x^i_j}}{2\beta_G} \sigma_G^2 
\]
for the Geometric Brownian Motion stem taper model
\[ 
\begin{aligned}
 p_{GB}(y^i_j, \Delta|x^i_j, \alpha_{GB}, \sigma_{GB}) &= \frac{1}{\sigma_{GB} \sqrt{2\pi x^i_j}} \exp \left( - \frac{\left( \ln \left( \frac{y^i_j}{y_0^i} \right) - \left( \alpha_{GB} - \frac{\sigma_{GB}^2}{2} \right) x^i_j \right)^2}{2\sigma_{GB}^2 x^i_j} \right) 
\end{aligned} 
\]
and for the Ornstein-Uhlenbeck stem taper model
\[ 
\begin{aligned}
 p_{OU}(y^i_j, \Delta|x^i_j, \alpha_{OU}, \beta_{OU}, \sigma_{OU}) &= \frac{1}{\sqrt{2\pi \sigma_{OU}(x^i_j)|\beta_{OU}, \sigma_{OU}} } \times \exp \left( - \frac{\left( y^i_j - m_{OU}(x^i_j|\alpha_{OU}, \beta_{OU}) \right)^2}{2\sigma_{OU}(x^i_j)|\beta_{OU}, \sigma_{OU}} \right) 
\end{aligned} 
\]
\[ 
\begin{aligned}
 m_{OU}(x^i_j|\alpha_{OU}, \beta_{OU}) &= y_0^i \exp \left( - \frac{x^i_j}{\beta_{OU}} \right) + \alpha_{OU} \beta_{OU} \left( 1 - \exp \left( - \frac{x^i_j}{\beta_{OU}} \right) \right) 
\end{aligned} 
\]
\[ 
\begin{aligned}
 \nu_{OU}(x^i_j|\beta_{OU}, \sigma_{OU}) &= \frac{\sigma_{OU}^2 \beta_{OU}}{2} \left( 1 - e^{-\frac{2x^i_j}{\beta_{OU}}} \right) 
\end{aligned} 
\]
The mean and variance functions \( m(x^i_j) \), \( \nu(x^i_j) \) (\( x^i_j \) is the relative height of \( i \)th tree) of the stochastic processes (1)-(3) are defined
\[ 
\begin{aligned}
 m_G(x^i_j|y^i_j, \alpha_G, \beta_G, \sigma_G) &= y_0^i e^{-\beta_G x^i_j} \\
 \times \frac{1 - e^{-\beta_G x^i_j}}{\beta_G} \left( \alpha_G - \frac{\sigma_G^2}{2} \left( 1 - e^{-\beta_G x^i_j} \right) \right) 
\end{aligned} 
\]
\[ 
\begin{aligned}
 v_G(x^i_j|y^i_j, \alpha_G, \beta_G, \sigma_G) &= \left( \ln \left( \frac{y^i_j}{y_0^i} \right) - \left( \alpha_G - \frac{\sigma_G^2}{2} \right) x^i_j \right)^2 \left( \alpha_G - \frac{\sigma_G^2}{2} \right) x^i_j \\
 + \frac{\sigma_G^2}{2\beta_G} \left( 1 - e^{-\beta_G x^i_j} \right) \right) \\
 \times \left( \exp \left( \frac{\sigma_G^2}{2\beta_G} \left( 1 - e^{-\beta_G x^i_j} \right) \right) - 1 \right) 
\end{aligned} 
\]
for the Gompertz model,
\[ 
\begin{aligned}
 v_{GB}(x^i_j|y^i_j, \alpha_{GB}, \sigma_{GB}) &= \left( \ln \left( \frac{y^i_j}{y_0^i} \right) - \left( \alpha_{GB} - \frac{\sigma_{GB}^2}{2} \right) x^i_j \right)^2 \left( \alpha_{GB} - \frac{\sigma_{GB}^2}{2} \right) x^i_j \\
 + \frac{\sigma_{GB}^2}{2\beta_{GB}} \left( 1 - e^{-\beta_{GB} x^i_j} \right) \right) \\
 \times \left( \exp \left( \frac{\sigma_{GB}^2}{2\beta_{GB}} \left( 1 - e^{-\beta_{GB} x^i_j} \right) \right) - 1 \right) 
\end{aligned} 
\]
for the Geometric Brownian Motion model, and for the Ornstein-Uhlenbeck model the mean and variance functions \( m(x^i_j) \), \( \nu(x^i_j) \) are defined by (9), (10).

Using transition densities (6)-(8) of SDEs (1)-(3), the transition probability density functions of relative diameter stochastic process \( Y^i_j(x^i_j) \), \( x^i_j \in [0, 1] \), \( i = 1, \ldots, M \), \( j = 0, \ldots, n_i \) defined by Eqs. (4), (5) take the forms (15), (16), respectively.

\[ 
\begin{aligned}
 p_G(y^i_j, x^i_j|y_0^i, \alpha_G, \beta_G, \sigma_G, \alpha_{GB}, \sigma_{GB}) &= \delta \left( y^i_j - y_0^i \right) \delta \left( x^i_j - x_0^i \right) \\
 p_{GB}(y^i_j, x^i_j|y_0^i, \alpha_{GB}, \sigma_{GB}) &= \delta \left( y^i_j - y_0^i \right) \delta \left( x^i_j - x_0^i \right) \\
 p_{OU}(y^i_j, x^i_j|y_0^i, \alpha_{OU}, \beta_{OU}, \sigma_{OU}) &= \delta \left( y^i_j - y_0^i \right) \delta \left( x^i_j - x_0^i \right) 
\end{aligned} 
\]
In this paper, we apply the theory of a one-stage maximum likelihood estimator for the stem taper SDE models (4), (5). Both models have closed form transition probability density functions (15), (16). Thus, the log-likelihood function for stem taper SDE models (4), (5) are

\[ LL_k(\alpha_G, \beta_G, \sigma_G, \alpha_{GB}, \sigma_{GB}, \alpha_{OU}, \sigma_{OU}) = \sum_{i=1}^{M} \sum_{j=1}^{n} \ln(p_y(y_j^i, y_j^i|\alpha_G, \beta_G, \sigma_G, \alpha_{GB}, \sigma_{GB}, \alpha_{OU}, \sigma_{OU})) \]

### Data

We focus on the modelling of Scots pine (Pinus Sylvestris) tree data sets. Scots pine trees dominate Lithuanian forests, growing on Arenosols and Podzols forest sites and covering 725500 ha. Stem measurements for 300 Scots pine trees were used for volume and stem profile models analysis. All section measurements include of 3821 data points. Summary statistics for diameter outside bark at breast height of 6.3 cm, 24.4 cm, 41.7 cm, and total tree heights of 6.8 m, 25.3 m, 33.1 m, respectively, are plotted in Fig. 1. Fig. 1 consists of stem taper curve (the mean of the stem taper) and standard deviation (the mean ± standard deviation of the stem taper). Fig. 1 illustrates the mean (Eq. (17), (19) – solid line) and standard deviation trajectory (Eq. (18), (20) – dash-dot line) of diameter at any given height. It is clear that these tree forms follow the stem data very closely. Graphical examination of residuals leads to the conclusion that SDEs stem taper models (3), (4) describe stem taper quite well.

### III. Results and Discussion

Estimation results are presented in Table 2. All parameters are highly significant (\( \alpha < 0.05 \)). The fit statistics (mean absolute prediction bias, least squares based Akaike’s [18] information criterion, and an adjusted coefficient of determination (\( R^2 \))) for the SDEs stem taper models (4), (5) are very close to the other commonly used stem taper regression models. The diameters' estimate along the stem for both SDEs stem taper models proves satisfactory, with mean absolute prediction bias 1.088 cm, 0.950 cm, respectively. The percent of variation explained attains high level 98.4%. To test the compatibility between taper and volume equations of all used stem taper models, the observed volume values from the sampled trees were calculated in the following form

\[ V_i = \frac{\pi}{40000} \left( \sum_{k=1}^{n-2} (d_{ik}^3 + d_{ik+1}^3 + d_{ik}^2 \cdot d_{ik+1} \cdot L_{ik} + \frac{d_{ik}^2 \cdot L_{ik} \cdot L_{ik+1}}{3}) \right) \]
Fig. 1. Tree profiles for three randomly selected pine trees generated using SDEs models (4), (5): (a) Eqs. (17), (18); (b) Eqs. (19), (20).

IV. CONCLUSIONS

Two new taper models were developed using stochastic differential equations. The both SDEs stem taper models provided here can be predicted from simple and standard tree measurements: total height, diameter outside bark at breast height, and diameter outside bark at the butt.

The SDEs approach allows us to incorporate new tree variables, mixed-effect parameters, and new forms of stochastic dynamics.

Despite the advantages of the SDEs stem taper models, it should be kept in mind that their main weakness is the sophisticated framework of the mathematical model.

Finally, we may notice that stochastic differential equation methodology may be of interest far beyond the modelling of a tree taper.

REFERENCES