



# Nonlinear parsimonious forest modeling assuming normal distribution of residuals

Bogdan M. Strimbu<sup>1</sup> · Alexandru Amarioarei<sup>2,3</sup> · Mihaela Paun<sup>3,4</sup>

Received: 25 March 2019 / Revised: 25 December 2020 / Accepted: 6 January 2021  
© The Author(s) 2021

## Abstract

To avoid the transformation of the dependent variable, which introduces bias when back-transformed, complex nonlinear forest models have the parameters estimated with heuristic techniques, which can supply erroneous values. The solution for accurate nonlinear models provided by Strimbu et al. (Ecosphere 8:e01945, 2017) for 11 functions (i.e., power, trigonometric, and hyperbolic) is not based on heuristics but could contain a Taylor series expansion. Therefore, the objectives of the present study are to present the unbiased estimates for variance following the transformation of the predicted variable and to identify an expansion of the Taylor series that does not induce numerical bias for mean and variance. We proved that the Taylor series expansion present in the unbiased expectation of mean and variance depends on the variance. We illustrated the new modeling approach on two problems, one at the ecosystem level, namely site productivity, and one at individual tree level, namely stem taper. The two models are unbiased, more parsimonious, and more precise than the existing less parsimonious models. This study focuses on research methods, which could be applied in similar studies of other species, ecosystem, as well as in behavioral sciences and econometrics.

**Keywords** Taylor series expansion · Unbiased estimates · Hyperbolic functions · Trigonometric functions · Power function · Mean · Variance

## Introduction

The formal departure from the linear modeling arguably starts with the development of derivatives of basic functions by Newton (1687) and von Leibnitz (1920). Nevertheless, the seminal work of Newton and Leibnitz on nonlinearity was implemented mainly on relatively simple formulations, such as trigonometric, power, or exponential functions.

Significant advancements in modeling occurred in 1715 when Taylor (1715) presented an approximation of any function that is locally any continuously differentiable with a polynomial function. Even after the introduction of the Taylor series, application to environmental processes was limited, until 1877, when Galton (1877) developed linear regression. One of the main advancements of the regression was the ability to represent nonlinearity by transforming the variables (Schumacher and Hall 1933; Warton and Hui 2011), in most instances the predictors (Neter et al. 1996). The linear regression coefficients were estimated for almost two hundred years with the least square method (Cotes 1722). However, even when the assumptions of the least squares method are met, the transformation of the predictors did not necessarily supply the desired results. In those cases, transformation of the predicted was executed to improve the results. However, the bias induced by the transformation of the dependent variable was formally addressed more than half century later by Williams (1937) and Cochran (1938). Nevertheless, bias correction for the case when the predicted variable was changed was developed only for few transformations, such as the logarithm function (Finney 1941).

---

Communicated by Arne Nothdurft.

---

Bogdan M. Strimbu  
bogdan.strimbu@oregonstate.edu

<sup>1</sup> College of Forestry, Oregon State University, Corvallis, OR 97331, USA

<sup>2</sup> Faculty of Mathematics and Computer Science, University of Bucharest, Bucharest, Romania

<sup>3</sup> Bioinformatics Department, National Institute of Research and Development for Biological Sciences, Bucharest, Romania

<sup>4</sup> Faculty of Administration and Business, University of Bucharest, Bucharest, Romania

Transformation of the dependent variable without correcting it was present even after the seminal paper of Neyman and Scott (1960), who proposed a bias-correction framework for almost all functions. The complex implementation of the Neyman and Scott framework (Neyman and Scott 1960) was overcome by the development of the generalized linear models (GLM) by Nelder and Wedderburn (1972). The assumptions of GLM limited its application to a reduced number functions, such as the logistic function.

The development in the information technology at the end of the second millennium, which allowed massive computations in a short amount of time, recommended new procedures for modeling complex nonlinear functions. For more than 50 years, the main estimators were nonlinear least squares, as proposed by Levenberg (1944) and Marquardt (1963), and the restricted maximum likelihood, as proposed Bartlett (1937) and formalized by Patterson and Thompson (1971). Both methods are suboptimal as either considered only a portion of the data, the case of restricted maximum likelihood, or do not search the entire solution space, the case of the nonlinear least squares. Therefore, new procedures were proposed, which are based on complex heuristics (Hoos and Stutzle 2005; Talbi 2009). The heuristic methods, such as simulated annealing, genetic algorithms, or particle swarm optimization, have the ability to either find the actual values of the parameters defining the nonlinear model or supply values close to the actual values in a reasonable amount of time (Aledo et al. 2016; Prieto-Escobar et al. 2018; Özsoy et al. 2020). A wave of developments of heuristic algorithms aiming at the estimation of parameters of nonlinear relationships happened at the beginning

of the third millennium, such as Pujol (2007), Yuan (2011, 2015), or Chen et al. (2008), to cite just a few. However, the sophistication of the heuristic techniques relies on the fact that an approximation of the solution is obtained. In many instances, the heuristic solutions are so close to the actual solution, that there is no practical reason to spend more effort in attaining better results (Bettinger et al. 2002). However, process-based modeling (Korzukhin et al. 1996) is sensitive to the solution supplied by the heuristic techniques, as incorrect relationships can alter fundamentally the behavior, and consequently the interpretation, of the ecosystem dynamic.

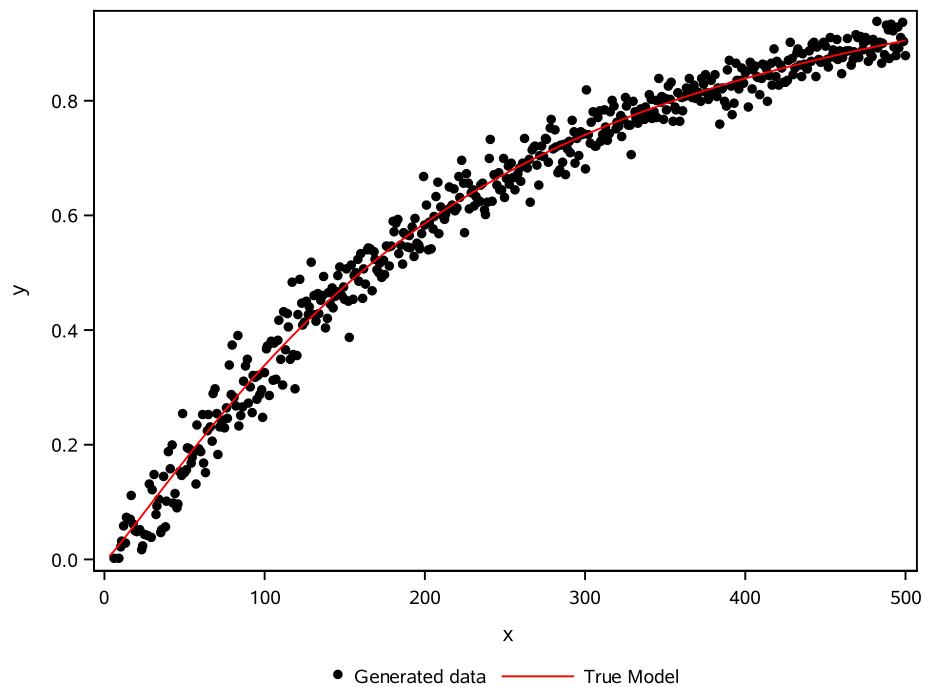
In many instances, common algorithms based on heuristics (e.g., steepest descent, Gauss–Newton, or Marquardt) estimate the parameters of nonlinear relationship with opposite signs than the actual ones. Bayesian approaches (Gelman et al. 2003) can lead to similar results, as proven by Amarioarei et al. (2020). To prove the impact of the estimation procedure on the parameters to be estimated, we use an example. Let assume that a process can be modeled with the equation:

$$y = \arctan \left( 5 - 5 \times \exp \left( \frac{\sqrt{x} - x}{1000} \right) \right) / 1.2 + e, \quad (1)$$

where  $x$  is the predictor,  $y$  the response variable, and  $\tan(y)$  at a given  $x$  is normally distributed with mean  $\mu_x$  and standard deviation of  $\sigma_x^2$ .

If synthetic  $y$  is generated for  $x$  varying from 1 to 500, assuming a normal distribution of the residuals of  $\tan(y)$  with variance 0.01 (Fig. 1), then the parameters of Eq. 1 can be estimated using the nonlinear model

**Fig. 1** Generated data based on Eq. 1



$$y = \arctan \left( b_0 + b_1 \times \exp \left( b_2 \sqrt{x} + b_3 x \right) \right) / b_4 + e. \quad (2)$$

The solution of Eq. 2 varies with the estimation algorithm, as the SAS implementation of the Marquardt algorithm supplies the values  $b_0 = -0.122$ ,  $b_1 = 0.0868$ ,  $b_2 = 0.2262$ ,  $b_3 = -0.00305$ , and  $b_4 = 1.3712$ , whereas the Gauss–Newton algorithm leads to  $b_0 = 2.4274$ ,  $b_1 = -2.3739$ ,  $b_2 = 0.0078$ ,  $b_3 = -0.0023$ , and  $b_4 = 1.0988$ . The difference between the computed values and the actual values is not necessarily important for this argument; what is important is the change in the sign of  $b_0$  and  $b_1$  between algorithms, which triggers a different model interpretation. The generated model has  $b_0$  positive and  $b_1$  negative, whereas the Marquardt algorithm supplies the opposite values for both and Gauss–Newton consistent with the sign of the model. Therefore, the heuristics employed in estimation of nonlinear models can supply incorrect results because of the algorithm. However, if the dependent variable is transformed using the tangent function, then the coefficient would have the correct sign but it would produce biased results when the predicted variable is back-transformed. Nevertheless, if correction of the bias induced by the nonlinear transformation of the dependent variable is applied, then the back-transformation would produce unbiased values.

The reduced number of nonlinear functions for which unbiased estimation exists (Nelder and Wedderburn 1972), the difficult to implement framework proposed by Neyman and Scott (1960) for bias correction when the dependent variable is transformed, and the lack of accuracy associated with heuristics prompted the development of unbiased estimates for 10 transformations of the predicted variable that are commonly encountered in forest modeling (Strimbu et al. 2017). For the 10 functions, complex nonlinear models can be developed exactly, as parameters are estimated without using heuristics. Furthermore, a sequential transformation of the dependent variable can now be applied, as the estimated values are accurate and precise.

The approach proposed by Strimbu et al. (2017), which avoids heuristic estimations, provides unbiased results when transforming the predicted variable,  $Y$ , with a differentiable function  $f$ . The method for correcting the bias induced by the change of the dependent variable is based on the assumption that between  $f(Y)$  and a set of predictor variables,  $X$ , there is a linear relationship

$$f(Y) = Xb + \varepsilon, \quad (3)$$

where  $b$  is the vector of coefficients for the independent variables  $X$ ,  $\varepsilon$  are the residuals, which are normally distributed with mean 0 and variance  $\sigma^2$ ,  $\varepsilon \sim N(0, \sigma^2)$ .

The bias correction based on Eq. 3 computed explicitly the mean of  $Y|X$  for 10 commonly used functions. Because

the estimates for eight of these functions (i.e., sine, cosine, tangent, arc sine, arc cosine, arc tangent, hyperbolic sine, and hyperbolic tangent functions) contains a Taylor series expansion, the formulas contain an infinite number of terms. The simulated data used by the author to guide the selection of the number of terms present in the Taylor series expansion is likely to be challenged by real problems, which are more complex than simulated data. Also, the method presented by Strimbu et al. (2017) does not provide unbiased estimates for the variance of the back-transformed  $Y$ . Estimation of variance of the predicted values is mandatory for computing the confidence intervals. Therefore, the objective of the present study is twofold: first, to present the unbiased estimates of the variance of the back-transformed variable, and second, to estimate a computational efficient Taylor series expansion that would provide unbiased results for the transformations involving Taylor series expansion. To illustrate the nonlinear estimation method advocated by this study, we present three forestry applications: one on site productivity, one on stem taper, and one on straw decomposition.

## Methods

### Foundation

Equation 3 can be rewritten  $f_1(Y) = Y_1 = Xb + \varepsilon$ , for which an unbiased estimation of  $Y$  given  $X=x$  is:

$$Y|_x = f_1^{-1}(Xb + \varepsilon)|_{x'}, \quad (4)$$

When another transformation is applied to the first transformation,  $f_2 \circ f_1$ , then Eq. 3 becomes

$$f_2(Y_1) = f_2(f_1(Y)) = X'b' + \varepsilon' \quad (5)$$

for which an unbiased estimator of  $Y$  at  $x'$  is according to Shanks and Gambill (1973):

$$Y|_{x'} = f_1^{-1}(f_2^{-1}(X'b' + \varepsilon'))|_{x'}, \quad (6)$$

where  $\varepsilon'$  is a normal distributed residual with mean 0 and variance  $\sigma'^2$ .

A series of unbiased back-transformations of nonlinear functions, each containing a finite number of terms, concludes with an unbiased estimate. Nevertheless, when an infinite number of terms are required but only few are used, the compounding from Eq. 5 could lead to computational errors (Amarioarei et al. 2020). Significant deviation from the actual model can occur because of the truncation, even when only one transformation is present, like Eq. 4. To assess the impact of truncation on the transformations, we focused on the functions that contains a Taylor series expansion, as computed by Strimbu et al. (2017). However, the

identification of an efficient selection of terms for the Taylor series expansion that would produce unbiased numerical estimates is not sufficient for modeling purposes. The variance of the predicted values is also required. In this study, we have also included formulas for the variance  $\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}^2[Y]$  (Grimmett and Stirzaker 2002), which requires the estimation of the mean,  $\mathbb{E}[Y]$ , and of the second moment,  $\mathbb{E}[Y^2]$ . The formulas for the expectation of  $Y$  were proven by (Strimbu et al. 2017); therefore, in this study we will include results for the  $\mathbb{E}[Y^2]$ . The final formulations for a series of common transformations are (proof in “Appendix”):

- Power:  $f(y) = y^a$  with  $a > 0$  and  $y \in (0, \infty)$

$$\begin{aligned}\mathbb{E}[Y] &= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \left( \frac{1}{a} \right) \xi^{\frac{1}{a}-k} \sigma^k \int_{-\frac{\xi}{\sigma}}^{\infty} t^k e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \left( \frac{1}{a} \right) \xi^{\frac{1}{a}-k} \sigma^k I(\alpha, k)\end{aligned}\quad (7)$$

$$\begin{aligned}\mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \left( \frac{2}{a} \right) \xi^{\frac{2}{a}-k} \sigma^k \int_{-\frac{\xi}{\sigma}}^{\infty} t^k e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \left( \frac{2}{a} \right) \xi^{\frac{2}{a}-k} \sigma^k I(\alpha, k)\end{aligned}\quad (8)$$

- Sine:  $f(y) = \sin(y)$  and  $y \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ , for which  $\alpha = -(1 + \xi)/\sigma$  and  $\beta = (1 - \xi)/\sigma$

$$\begin{aligned}\mathbb{E}[Y] &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(2n-1)!!}{(2n)!!} \frac{1}{2n+1} \\ &\quad \sum_{k=0}^{2n+1} \left( \begin{array}{c} 2n+1 \\ k \end{array} \right) \xi^{2n+1-k} \sigma^k I(\alpha, \beta, k)\end{aligned}\quad (9)$$

$$\begin{aligned}\mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{4^n (n!)^2}{(2n+1)!(n+1)} \\ &\quad \sum_{k=0}^{2n+2} \left( \begin{array}{c} 2n+2 \\ k \end{array} \right) \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k)\end{aligned}\quad (10)$$

- Cosine:  $f(y) = \cos(y)$  and  $y \in [0, \pi]$ , for which  $\alpha = -(1 + \xi)/\sigma$  and  $\beta = (1 - \xi)/\sigma$

$$\begin{aligned}\mathbb{E}[Y] &= \frac{\pi}{2} (\Phi(\beta) - \Phi(\alpha)) - \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \\ &\quad \frac{(2n-1)!!}{(2n)!!} \frac{1}{2n+1} \sum_{k=0}^{2n+1} \left( \begin{array}{c} 2n+1 \\ k \end{array} \right) \xi^{2n+1-k} \sigma^k I(\alpha, \beta, k)\end{aligned}\quad (11)$$

$$\begin{aligned}\mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \left[ -\frac{(2n-1)!!}{(2n)} \frac{\pi}{2n+1} \sum_{k=0}^{2n+1} \left( \begin{array}{c} 2n+1 \\ k \end{array} \right) \xi^{2n+1-k} \sigma^k I(\alpha, \beta, k) \right. \\ &\quad \left. + \frac{4^n (n!)^2}{(2n+1)!(n+1)} \sum_{k=0}^{2n+2} \left( \begin{array}{c} 2n+2 \\ k \end{array} \right) \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k) \right] + \frac{\pi^2}{4} (\Phi(\beta) - \Phi(\alpha))\end{aligned}\quad (12)$$

- Tangent:  $f(y) = \tan(y)$  and  $y \in \left[0, \frac{\pi}{4}\right]$ , for which  $\alpha = -\xi/\sigma$  and  $\beta = (1 - \xi)/\sigma$

$$\mathbb{E}[Y] = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \sum_{k=0}^{2n+1} \left( \begin{array}{c} 2n+1 \\ k \end{array} \right) \xi^{2n+1-k} \sigma^k I(\alpha, \beta, k)\quad (13)$$

$$\mathbb{E}[Y^2] = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n (2H_{2n+1} - H_n)}{2n+1} \sum_{k=0}^{2n+2} \left( \begin{array}{c} 2n+2 \\ k \end{array} \right) \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k),\quad (14)$$

where  $H_n = \sum_{k=0}^n \frac{1}{k}$  is the  $n$ th harmonic number.

- Arcsine:  $f(y) = \arcsin(y)$  and  $y \in [-1, 1]$ , for which  $\alpha = -(\pi/2 + \xi)/\sigma$  and  $\beta = (\pi/2 - \xi)/\sigma$

$$\mathbb{E}[Y] = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} \sum_{k=0}^{2n+1} \left( \begin{array}{c} 2n+1 \\ k \end{array} \right) \xi^{2n+1-k} \sigma^k I(\alpha, \beta, k)\quad (15)$$

$$\mathbb{E}[Y^2] = \frac{1}{2} \left[ \Phi(\beta) - \Phi(\alpha) - \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n 4^n}{(2n)!} \sum_{k=0}^{2n} \left( \begin{array}{c} 2n \\ k \end{array} \right) \xi^{2n-k} \sigma^k I(\alpha, \beta, k) \right]\quad (16)$$

- Arccosine:  $f(y) = \arccos(y)$  and  $y \in [-1, 1]$ , for which  $\alpha = -\xi/\sigma$  and  $\beta = (\pi - \xi)/\sigma$

$$\mathbb{E}[Y] = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} \sum_{k=0}^{2n} \left( \begin{array}{c} 2n \\ k \end{array} \right) \xi^{2n-k} \sigma^k I(\alpha, \beta, k)\quad (17)$$

$$\mathbb{E}[Y^2] = \frac{1}{2} \left[ \Phi(\beta) - \Phi(\alpha) + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n 4^n}{(2n)!} \sum_{k=0}^{2n} \left( \begin{array}{c} 2n \\ k \end{array} \right) \xi^{2n-k} \sigma^k I(\alpha, \beta, k) \right]\quad (18)$$

- Arctangent:  $f(y) = \arctan(y)$  with  $y \in \left[0, \frac{\pi}{4}\right]$ , for which  $\alpha = \frac{-\frac{\pi}{4} - \xi}{\sigma}$  and  $\beta = \frac{\frac{\pi}{4} - \xi}{\sigma}$

$$\mathbb{E}[Y] = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n 2^{2n} (2^{2n} - 1) B_{2n}}{(2n)!} \sum_{k=0}^{2n-1} \left( \begin{array}{c} 2n-1 \\ k \end{array} \right) \xi^{2n-1-k} \sigma^k I(\alpha, \beta, k),\quad (19)$$

where  $B_n$  is the  $n$ th Bernoulli number, computed recursively starting from  $n=0$  and  $B_0=1$  with the formula  $B_n = - \sum_{k=0}^{n-1} \left( \begin{array}{c} n \\ k \end{array} \right) \frac{B_k}{n-k-1}$ ,  $n \geq 1$ .

- Hyperbolic sine:  $f(y) = \sinh(y) = \frac{e^y - e^{-y}}{2}$  with  $y \in [0, a]$ , for which  $\alpha = -\xi/\sigma$  and  $\beta = (1 - \xi)/\sigma$

$$\mathbb{E}[Y] = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n (2n)!}{2^{2n} (n!)^2 (2n+1)} \sum_{k=0}^{2n+1} \binom{2n+1}{k} \xi^{2n+1-k} \sigma^k I(\alpha, \beta, k) \quad (20)$$

$$\mathbb{E}[Y^2] = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-4)^n (n!)^2}{(n+1)(2n+1)!} \sum_{k=0}^{2n+2} \binom{2n+2}{k} \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k) \quad (21)$$

- Hyperbolic arcsine:  $f(y) = \text{arcsinh}(y) = \ln(y + \sqrt{y^2 + 1})$ ,  $y \in [0, \infty)$

$$\mathbb{E}[Y] = \frac{1}{2} e^{\frac{\sigma^2}{2}} \left[ 2 \sinh(\xi) + e^{-\xi} \Phi\left(-\frac{\xi}{\sigma} - \sigma\right) - e^{\xi} \Phi\left(-\frac{\xi}{\sigma} + \sigma\right) \right] \quad (22)$$

$$\begin{aligned} \mathbb{E}[Y^2] = & -\frac{1}{2} \left[ 1 - \Phi\left(-\frac{\xi}{\sigma}\right) \right] \\ & + \frac{e^{2\sigma^2}}{2} \left[ 2 \cosh(2\xi) - e^{2\xi} \Phi\left(-\frac{\xi}{\sigma} - 2\sigma\right) - e^{-2\xi} \Phi\left(-\frac{\xi}{\sigma} + 2\sigma\right) \right]. \end{aligned} \quad (23)$$

- Hyperbolic tangent:  $f(y) = \tanh(y) = \frac{e^y - e^{-y}}{e^y + e^{-y}}$  with  $y \in [0, \infty)$ , for which  $\alpha = -\xi/\sigma$  and  $\beta = (1 - \xi)/\sigma$

$$\mathbb{E}[Y] = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sum_{k=0}^{2n+1} \binom{2n+1}{k} \xi^{2n+1-k} \sigma^k I(\alpha, \beta, k) \quad (24)$$

$$\mathbb{E}[Y^2] = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{2H_{2n+1} - H_n}{2n+1} \sum_{k=0}^{2n+2} \binom{2n+2}{k} \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k) \quad (25)$$

with  $H_n$  being the  $n$ th harmonic number.

The symbols used in Eqs. 9–25 are:

$n, k$  are natural numbers;

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$  is the binomial coefficient;

$(2n-1)!! = 1 \times 3 \times \dots \times (2n-1)$  and

$(2n)!! = 2 \times 4 \times \dots \times (2n)$  are the products of the uneven, or even, positive integers  $\leq 2n-1$  or  $2n$ ;

$\xi = x^T b$ , with  $x^T$  being the transpose of the vector  $x$ ;  $(26)$

$$I(\alpha, \beta, n) = \int_{\alpha}^{\beta} t^n e^{-\frac{t^2}{2}} dt \text{ and } I(\alpha, n) = \lim_{\beta \rightarrow \infty} I(\alpha, \beta, n). \quad (27)$$

Denoting by  $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ , the standard normal density function and by  $\Phi(x) = \int_{-\infty}^x \phi(t) dt$  the standard normal cumulative distribution function, it can be shown (for example, by mathematical induction) that the integral  $I(\alpha, \beta, n)$  is given by

$$\begin{cases} \sqrt{2\pi}(n-1)!! \left[ \phi(\alpha) \left( 1 + \sum_{k=0}^{\frac{n-1}{2}-1} \frac{\alpha^{2k+2}}{(2k+2)!!} \right) - \phi(\beta) \left( 1 + \sum_{k=0}^{\frac{n-1}{2}-1} \frac{\beta^{2k+2}}{(2k+2)!!} \right) \right], & n \text{ odd} \\ \sqrt{2\pi}(n-1)!! \left[ \Phi(\beta) - \Phi(\alpha) + \phi(\alpha) \sum_{k=0}^{\frac{n}{2}-1} \frac{\alpha^{2k+1}}{(2k+1)!!} - \phi(\beta) \sum_{k=0}^{\frac{n}{2}-1} \frac{\beta^{2k+1}}{(2k+1)!!} \right], & n \text{ even} \end{cases} \quad (28)$$

with initial values  $I(\alpha, \beta, 0) = \sqrt{2\pi}(\Phi(\beta) - \Phi(\alpha))$  and  $I(\alpha, \beta, 1) = \sqrt{2\pi}(\phi(\alpha) - \phi(\beta))$ .

For  $\beta \rightarrow \infty$ , the integral from Eq. 28,  $\lim_{\beta \rightarrow \infty} I(\alpha, \beta, n) = I(\alpha, n)$ , simplifies to:

$$I(\alpha, n) = \begin{cases} \sqrt{2\pi} \phi(\alpha) (n-1)!! \left[ 1 + \sum_{k=0}^{\frac{n-1}{2}-1} \frac{\alpha^{2k+2}}{(2k+2)!!} \right], & \text{for odd} \\ \sqrt{2\pi} (n-1)!! \left[ 1 - \Phi(\alpha) + \phi(\alpha) \sum_{k=0}^{\frac{n}{2}-1} \frac{\alpha^{2k+1}}{(2k+1)!!} \right], & \text{for even} \end{cases} \quad (29)$$

where  $I(\alpha, 0) = \sqrt{2\pi}(1 - \Phi(\alpha))$  and  $I(\alpha, 1) = \sqrt{2\pi}\phi(\alpha)$ .

We considered an efficient Taylor series expansion a series for which the addition of a new term fulfills two conditions: first, the improvement of the existing estimates is  $< 10^{-5}$ , and second, the line constructed from estimates computed from two consecutive number of terms is almost horizontal (i.e., the slope is  $< 2^\circ$ ). The conditions mirror the two common approaches used for selection of terms from a series: one based on a preset value (LeVeque 2007), and one based on the scree method (Cattell 1966; Tabachnick and Fidell 2001; Zhu and Ghodsi 2006).

By truncation, the expression of the means from Eqs. 9–25 can be rewritten in matrix form such that the Taylor series expansion would include only the first  $N$  terms:

$$\hat{y} = \sum_{k=0}^{2N+1} \sum_{n=0}^N c_{k,n} \Xi_{n,k} = \text{Tr}(\mathbf{C}\Xi), \quad (30)$$

where  $\hat{y}$  is the predicted, back-transformed variable

$$\mathbf{C} = (c_{k+1,n+1})_{\substack{0 \leq k \leq 2N+1 \\ 0 \leq n \leq N}} \in \mathcal{M}_{2N+2, N+1}, \quad (31)$$

$$\Xi = (\Xi_{n+1,k+1})_{\substack{0 \leq n \leq N \\ 0 \leq k \leq 2N+1}} \in \mathcal{M}_{N+1, 2N+2} \quad (32)$$

$$\Xi_{n,k} = \begin{cases} \xi^{2n-k} \sigma^k I(\alpha, \beta, k) \mathbf{1}_{\{0, \dots, 2n\}}(k), & \text{arccosine} \\ \xi^{2n-1-k} \sigma^k I(\alpha, \beta, k) \mathbf{1}_{\{0, \dots, 2n-1\}}(k), & \text{arctangent, for } 0 \leq n \leq N. \\ \xi^{2n+1-k} \sigma^k I(\alpha, \beta, k) \mathbf{1}_{\{0, \dots, 2n+1\}}(k), & \text{other} \end{cases} \quad (33)$$

The terms  $c_{k,n}$  from the triangular matrix  $\mathbf{C}$  depend on the transformation, and are given by

$$c_{k,n} = \begin{cases} \frac{(2n-1)!!}{(2n)!!} \frac{1}{2n+1} \binom{2n+1}{k}, & \text{sine, cosine} \\ \frac{(-1)^n}{(2n+1)!} \binom{2n+1}{k}, & \text{arcsine} \\ \frac{(-1)^n}{(2n)!} \binom{2n}{k}, & \text{arccosine} \\ \frac{(-1)^{n-1} 2^{2n} (2^{2n}-1) B_{2n}}{(2n)!} \binom{2n-1}{k}, & \text{tangent} \\ \frac{(-1)^n}{2n+1} \binom{2n+1}{k}, & \text{arctangent} \\ \frac{(-1)^n (2n)!}{2^{2n} (n!)^2 (2n+1)} \binom{2n+1}{k}, & \text{hyperbolic sine} \\ \frac{1}{2n+1} \binom{2n+1}{k}, & \text{hyperbolic tangent} \end{cases} \quad (34)$$

The elements of the matrix  $\Xi, \Xi_{n,k}$  depend on the variance of the linear model. Therefore, the selection of the significant terms triggers the back-transformation of the predicted variable as the product of the constant matrix  $\mathbf{C}$  and the matrix  $\Xi$ , which contains the linear regression values. It should be noticed that the matrix  $\mathbf{C}$  is completely defined by the transformation and the number of terms selected from the Taylor series expansion and are independent of the data for which the model is developed.

Considering that the first and second moments from Eqs. 9–25 depend on the variance, Amarioarei et al (2020) used a factorial experiment to prove that relatively few number of terms are required to represent the expectation unbiased (i.e., less than 10). However, real applications revealed that more than 10 terms could be needed for unbiased estimates, likely 20 or even 30. However, the large number of terms could be associated with the implementation, as computational algorithms were proven to play a significant role in the results (Seppelt and Richter 2005; Paun et al. 2020).

### Stand-level models for height of dominant Norway spruce (*Picea abies* L.)

The main economic species in the Carpathian Mountains is Norway spruce (Tudoran and Zotta 2020), which triggered significant efforts in modeling height of dominant trees. A polymorphic model with six parameters for height was developed by Giurgiu and Draghiciu (2004), based on the field measurements published in 1957 (Popescu-Zeletin 1957). The polymorphic model is enforced by the Romanian regulations, and consequently with significant impact on forest management, is

$$\text{height}_{\text{dominant}} = a_0 \times \text{SI} \times \left( t_{\text{Lorey height}} + t_{\text{Lorey height}}^{a_1} \times e^{a_2 \times t_{\text{Lorey height}}} \right), \quad (35)$$

where  $\text{height}_{\text{dominant}}$  is the height of dominant trees,  $t_{\text{Lorey height}}$  is transformed Lorey's height modeled with the expression  $e^{b_0(\text{age}^{b_1} - 100^{b_1})}$ ,  $a_0, a_1, a_2, b_0, b_1, b_2$  are parameters determined by species, SI is site index.

The models proposed by Giurgiu and Draghiciu (2004) are not justifiable, as they identify a regression with a causal relationships, which was proven to be false (Neter et al. 1996; Duursma and Robinson 2003). The biased models of Giurgiu and Draghiciu were replaced in 2020 by Amarioarei et al (2020), who developed parsimonious unbiased models. Using the same data as Amarioarei et al (2020), we improved the existing models not only in terms of estimators but also by supplying estimates for variance. Evermore, we have expanded the existing models by including the most productive sites, the ones labeled class I, which were not included in the original study of Amarioarei et al. Therefore, we proposed a new set of polymorphic equations for the height of dominant and codominant trees based on the hyperbolic tangent of the ratio between height and the preset site index, SI:

$$f(\text{height}) = \tanh \left( \frac{\text{height}}{\text{SI}} \right) = \frac{\exp \left( \frac{\text{height}}{\text{SI}} \right) - \exp \left( -\frac{\text{height}}{\text{SI}} \right)}{\exp \left( \frac{\text{height}}{\text{SI}} \right) + \exp \left( -\frac{\text{height}}{\text{SI}} \right)}, \quad (36)$$

where SI is height at age 100 years, which is 36.9 m for class I, 31.8 m for class II, 26.9 m for class III, and 21.9 m for class IV.

Amarioarei et al. found that the reciprocal of  $\sqrt{\text{age}}$  is linearly related to  $\tanh \left( \frac{\text{height}}{\text{SI}} \right)$ . Therefore, the height model is:

$$f(\text{height}) = b_0 + b_1 \times \frac{1}{\sqrt{\text{age}}} + e, \quad (37)$$

where  $e \sim N(0, \sigma_{\text{height}}^2)$ .

The number of terms needed to avoid numerical bias was identified by considering 5, 10, 20 and 30 terms in the Taylor series expansion. To assess the parsimonious model of height yield represented by Eq. 37, we compared it with the models currently used by the Romanian Forest Administration (Giurgiu and Draghiciu 2004), namely Eq. 35. To prove that the models obtained with the proposed study are not only unbiased but also efficient, we have supplemented the parametric estimation with the Bayesian approach proposed by Stow et al. (2006) to correct the retransformation bias. The Bayesian approach to nonlinear modeling is not new and was advocated by many studies as an alternative to the parametric estimates (van Oijen 2017; Golivets et al. 2019; Kansanen et al. 2019). We used R (Gentleman and Ihaka 2014) implementation of Stan (Stan Development Team 2016a) for Bayesian inference, namely the rstan package

(Stan Development Team 2016b). The Bayesian estimates produced by rstan, which is based on Markov Chain Monte Carlo, require a series of parameters, out of which the number of iterations is the most important. We followed the recommendation of the Stan Development Team (2016b), which suggested four chains. The lack of knowledge on data distribution, recommended the usage of a non-informative prior distribution in the computations, as suggested by Stow et al. (2006).

### Stem taper models for loblolly pine (*Pinus taeda* L.) in east Louisiana

Loblolly pine (*Pinus taeda* L.), which is one of the most important commercial tree species in the USA, was the subject of many stem taper models (Max and Burkhardt 1976; Cao et al. 1980; Fang et al. 2000; McClure and Czaplewski 2011; Fang and Strimbu 2017; Nicoletti et al. 2020). The majority of the stem taper models have at least two parameters, an exception being the model developed by Lenhart et al. (1987) who has one parameter. Amarioarei et al. (2020) proposed a trigonometric model for describing the stem taper of loblolly pine using the data from Fang and Strimbu (2017). The dataset contains 18 trees from the Vernon Parish, Louisiana, with diameters measuring every meter along the stem, plus the diameter at 1.3 m, breast height (dbh). We found that the cosine of the ratio between diameter,  $d$ , and double dbh (i.e.,  $\cos\left(\frac{d}{2 \times \text{dbh}}\right)$ ) is linearly related with the ratio of total height and square root of dbh and the product of the logarithmized total height [i.e.,  $\ln(\text{Total height})$ ] and the square root of relative height (i.e., height of diameter  $d$ /total height):

$$f(d) = \cos\left(\frac{d}{2 \times \text{dbh}}\right) = b_0 + b_1 \times \ln(\text{Total height}) \\ \times \sqrt{\frac{\text{height}}{\text{Total height}}} + b_2 \times \frac{\text{Total height}}{\sqrt{\text{dbh}}} + e, \quad (38)$$

where  $e \sim N(0, \sigma^2_d)$ .

Considering that the transformation of the predicted variable is cosine, the unbiased diameter along the stem was computed with Eq. 11, and the unbiased variance as the difference between Eqs. 12 and 11. We identified the number of terms needed to avoid numerical biased by considering 5, 10, 20, and 30 terms in the Taylor series expansion.

The model 02 of Kozak (2004) was demonstrated by Fang and Strimbu (2017) to be the most suitable to the 18 trees used in the present study. Therefore, we compared the parsimonious model obtained with Eq. 38 with Kozak02 model. To ensure consistency of the results, we determined the Bayesian estimates of the model presented by Eq. 38, using Stan's probabilistic language (Stan Development

Team 2016a) as implemented in the R programming language (Gentleman and Ihaka 2014). Similarly to the site index models, we have used four Markov chains and a non-informative prior distribution in computations. Stem taper models are subject to autocorrelation (Lindstrom and Bates 1990), which violates the independence assumption (Valentine and Gregoire 2001). However, when the residuals are white noise, then the models are considered complete (Brockwell and Davis 1996). Therefore, we tested autocorrelation with the Durbin–Watson test, as implemented in SAS 9.4 (Ansley et al. 1992). To ensure that the models developed are efficient, we have used Bartlett's test to assess homoskedasticity.

### Models assessment

The key requirement of the proposed nonlinear modeling framework mandatory represented by Eqs. 9–25 is the normal distribution of the residuals produced from the linear relationship between the transformed variable of interest and the independent variables. To ensure that the normality condition is fulfilled, we have used the Shapiro–Wilk test, as implemented in the base R programming language. Ensuing normal distribution of the residual, the performance of the three models were assessed using four metrics: pseudo- $R^2$ , bias, mean absolute error (MAE), and root mean square error (RMSE), similarly to Bilskie and Hagen (2013), Montealegre et al. (2015), and Stångle et al. (2017):

$$\text{bias} = \sum_{i=1}^n \frac{\hat{y}_i - y_i}{n} \quad (39)$$

$$\text{MAE} = \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{n} \quad (40)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}, \quad (41)$$

where  $\hat{y}_i, y_i, \bar{y}_i$  is the predicted, measured, or mean value at moment  $i$  (i.e., age, height, day),  $n$  is the number of observations.

All the computations were executed in R programming language (Gentleman and Ihaka 2014). We considered that an efficient Taylor series expansion is a series for which  $MAE < 1\%$ , similarly to Amarioarei et al. (2020).

## Results

### Foundation

The factorial experiment focused on Taylor series expansion showed that an efficient Taylor series could have  $< 10$  terms, regardless of the variance. The number of terms depends on the selection approach, with the preset value supplying more terms than the scree plot. However, the two approaches agreed, when the number of terms is chosen first time when the preset value is reached by the difference between the expectations computed with consecutive number of terms (i.e., one term apart), except for the hyperbolic tangent case (i.e., two terms apart). When the preset value was surpassed at least two times, the scree plot and the preset value approaches could differ by six terms (Amarioarei et al. 2020). Considering that the expectation of the first-order moment depends on the variance, the usage of a fewer terms should be avoided, as it can lead to biased results. To ensure computational unbiasedness, we considered that the number of terms should be selected according to the preset value, the case when it was met at least two times.

The second-order moment, and consequently the variance, depend on the same three estimates as the first-order moment, namely  $\xi$ ,  $\sigma$ , and  $I(\alpha, \beta, n)$ . Evermore, the formula for the first- and second-order moments is similar, the main difference being in the power of  $\xi$  or  $\sigma$  (Eqs. 7–25). Therefore, not surprisingly, the number of terms needed for unbiased estimates of the variance is also in the same range with the first-order moment, with the observation that usually an extra term is present. Therefore, a Taylor series expansion with 10 terms will likely ensure the lack of computational bias for the expectations, the predicted values, and the variance. Nevertheless, in complex applications, with larger variance there is the possibility that a larger number of terms would be needed in the Taylor series expansion.

### Models of dominant height of Romanian Norway spruce

According to Eq. 24, the height of dominant and codominant trees for the Romanian Norway spruce developed from Eq. 37 is

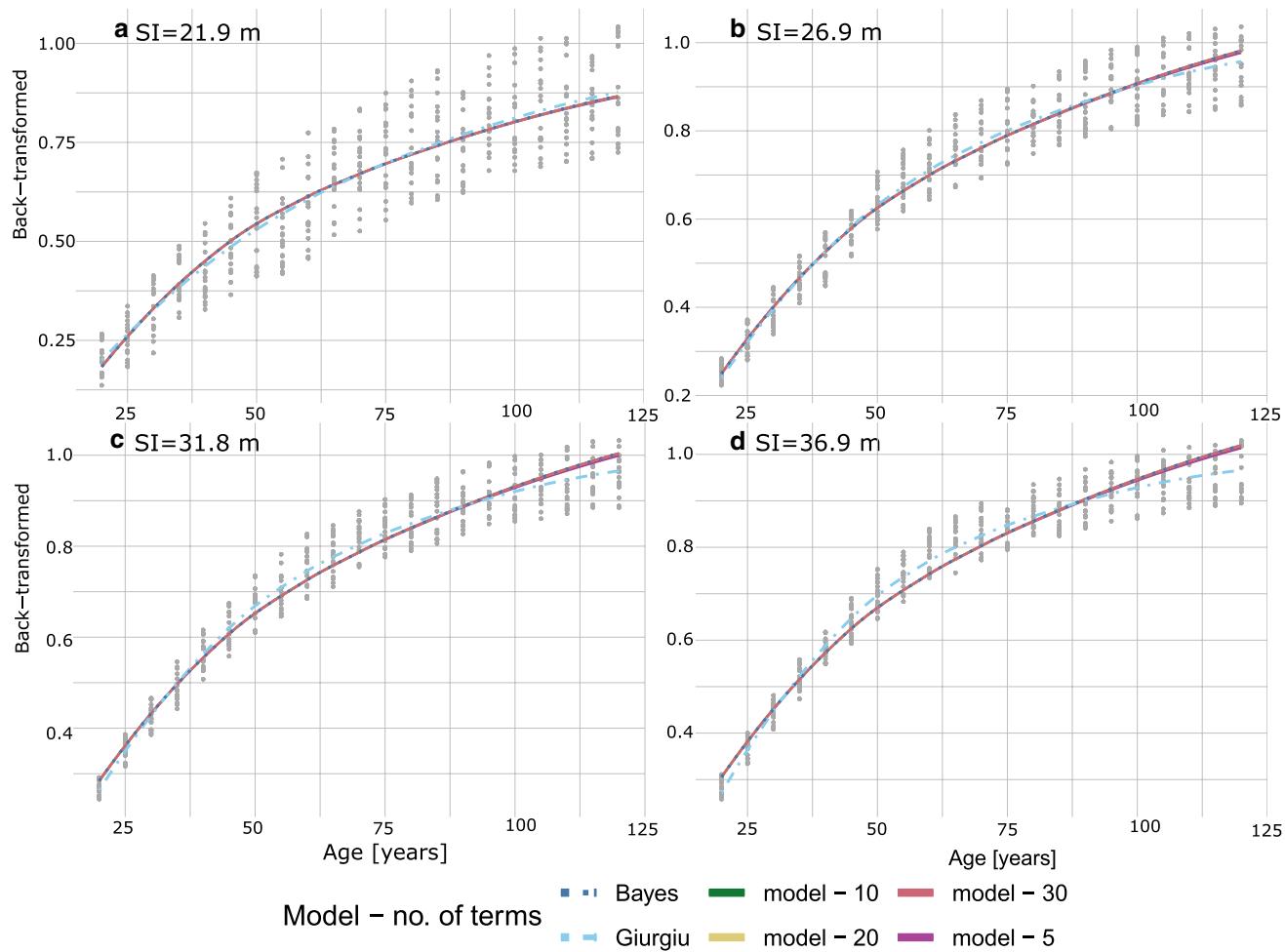
$$\widehat{\text{height}} = \sum_{k=0}^{2N+1} \sum_{n=0}^N \frac{1}{2n+1} \binom{2n+1}{k} \left( b_0 - \frac{b_1}{\sqrt{\text{age}}} \right)^{2n+1-k} \sigma^k I(\alpha, \beta, k) \mathbf{1}_{\{0, \dots, 2n+1\}}(k). \quad (42)$$

The coefficients  $b_0$  and  $b_1$  vary with site index (Table 1), which suggests that the height of dominant and codominant Norway spruce should be modeled with polymorphic equations.

All the proposed models are significant ( $p$  value  $< 0.001$ ) and have coefficient of determination  $> 0.9$ , which support their suitability in modeling the total height of dominant and codominant Norway spruce from Romania (Fig. 2). The validity of the models is supported by the Shapiro–Wilk test, which suggests normality of the residuals ( $p$  value  $> 0.25$ ); consequently, the appropriateness of Eq. 37 in modeling tree height. As expected, the correlation between the predicted and original data increases with the number of terms (Table 2), but minutely (i.e.,  $< 1\%$  from 10 to 30 terms). The Bayesian model based on Eq. 37 is almost identical with the one from Eq. 42, therefore the assessment metrics were similar (Table 2); with bias,  $MAE$  and  $RMSE$  being almost undetectable larger than the back-transformed values. The model currently in used in Romania (Giurgiu and Draghiciu 2004) had not surprisingly all six variables significant ( $p$  value  $< 0.001$ ). The correlation coefficient between the predicted and original data is superior to the one supplied by Eq. 42 (Table 2), but the difference is  $< 1\%$  for all site indices. Regardless of the site index, the bias,  $MAE$ , and  $RMSE$  were the smallest for Giurgiu and Draghiciu (2004) models (Table 2), but the differences were minuscule ( $< 1\%$ ). If the values predicted by the Giurgiu and Draghiciu (2004) models were infinitesimal superior to the Bayesian and the proposed methods, the situation is completely reversed when the focus is on variance. Irrespective of the assessment metric, the Bayesian and parsimonious models always exhibit a smaller variance than Giurgiu and Draghiciu (2004) models (Table 2). For bias, the variance of parsimonious models is three times smaller than the one supplied by the Bayesian or Giurgiu and Draghiciu (2004) models, sometimes almost one order of magnitude (e.g., 95% vs  $-9.7$  for  $SI = 26.9$  m). The same conclusion is reached for  $MAE$  and  $RMSE$ , with the parsimonious models consistently providing smaller variance, sometimes almost three times smaller (i.e., 95% vs 29% for the  $MAE$  of  $SI = 26.9$  m). Considering that the Giurgiu and Draghiciu (2004) models are less parsimonious than Eq. 42, exhibits significantly larger variances of the three metrics (inferior precision), while revealing an infinitesimal superior accuracy (i.e.,  $< 1\%$ ), it can be inferred that they are no longer justified in applications.

**Table 1** Coefficient of the height of dominant and codominant tree height according to Eq. 42

Coefficient	SI (m)			
	21.9	26.9	31.8	36.9
$b_0$	1.107664	1.116752	1.10624	1.098149
$b_1$	-4.3304	-3.9883	-3.75759	-3.62727



**Fig. 2** Model for the dominant height of Norway spruce in the Carpathian Mountains: **a** class IV ( $SI=21.9$  m), **b** class III ( $SI=26.9$  m), **c** class II ( $SI=31.8$  m), and **d** class I ( $SI=36.9$  m)

**Table 2** Assessment of the dominant/codominant tree height models for the Norway spruce in the Carpathian Mountains. The abbreviations are: Giurgiu Giurgiu–Draghiciu, Pars. Parsimonious, Var variance

Site index	Model	Bias (%)		MAE (%)		RMSE (%)		Correlation coefficient (%)
		Mean	Var	Mean	Var	Mean	Var	
21.9	Giurgiu	-2.1	84	12.9	84	8.7	84	91.7
	Bayesian	-2.0	82	13.3	82	8.8	83	91.4
	Pars. 10 terms	-1.9	12	13.3	50	8.8	57	91.4
	Pars. 30 terms	-2.0	12	13.3	50	8.8	57	91.4
26.9	Giurgiu	0.01	95	6.1	95	7.3	95	97.4
	Bayesian	-0.05	94	6.3	94	7.7	94	97.1
	Pars. 10 terms	-0.04	9.7	6.3	29	7.7	72	97.1
	Pars. 30 terms	-0.04	9.7	6.3	29	7.7	72	97.1
31.8	Giurgiu	3.4	96.3	4.8	96.3	5.6	96.3	98.1
	Bayesian	-0.5	95.3	5.1	95.3	6.0	95.3	97.7
	Pars. 10 terms	-0.5	-25.9	5.1	72.3	6.0	84.1	97.7
	Pars. 30 terms	-0.5	-25.9	5.1	72.3	6.0	84.1	97.7
36.9	Giurgiu	-0.08	96.8	4.3	96.8	5.2	96.8	98.3
	Bayesian	-0.5	94.9	4.9	94.9	6.1	94.9	97.4
	Pars. 10 terms	-0.5	-38.7	4.9	80.4	6.1	94.6	97.4
	Pars. 30 terms	-0.5	-38.7	4.9	80.4	6.1	94.6	97.4

## Stem taper models for loblolly pine

Equation 38 that modeled the stem taper model for loblolly pine is

$$\begin{aligned} \frac{\hat{d}}{2\text{dbh}} &= \sum_{k=0}^{2N+1} \sum_{n=0}^N \frac{1}{2n+1} \binom{2n+1}{k} \\ &\left( 0.783 + 0.0318 \times \log(\text{Total height}) \times \sqrt{\frac{h}{\text{Total height}}} \right. \\ &\left. + 0.01867 \times \frac{\text{Total height}}{\sqrt{\text{dbh}}} \right)^{2n+1-k} \\ &\sigma^k I(\alpha, \beta, k) \mathbf{1}_{\{0, \dots, 2n+1\}}(k), \end{aligned} \quad (43)$$

where  $\hat{d}$  is the predicted diameter measured at height  $h$ .

Equation 38 has both terms significant ( $p$  value  $< 0.0001$ ) and supplied a correlation coefficient between the predicted and the original data of  $-0.96$ . The normality test did not suggest a lack of normal distribution of the residuals ( $p$  value  $= 0.25$ ), which provides evidence that the parsimonious model is appropriate for stem taper modeling. The coefficient of correlation between the predicted and original data is virtually the same when the Taylor series expansion contains 10 or 30 terms, namely 0.96 (Table 3). Mirroring the height modeling, the Bayesian estimates were similar to the parsimonious model only in terms of bias but not in terms of variability, which was always two times larger (Table 3).

The nine parameters model of from Kozak (2004) was the most appropriate to describe the stem taper (Eq. 44), with a correlation coefficient between the predicted and original data of more than 98%. However, even that the model was significant as a whole ( $p$  value  $< 0.0001$ ), some of the variables were deemed insignificant, such as  $b_2$ , the coefficient of the inverse of the slenderness coefficient, or  $b_3$ , the coefficient of  $X^{0.1}$  ( $p$  values  $> 0.26$ ):

$$\begin{aligned} \frac{\hat{d}}{2\text{dbh}} &= 17.7256\text{dbh}^{0.573}(\text{Total height})^{-1.832} X^{g(\text{dbh}, h, \text{Total height}, X, Q)} \\ &\text{where } g(\text{dbh}, h, \text{Total height}, X, Q) \\ &= -0.394 \left( \frac{h}{\text{Total height}} \right)^4 - 0.08e^{-\frac{\text{dbh}}{\text{Total height}}} \\ &\quad + 0.056X^{0.1} + \frac{2.7}{\text{dbh}} + 0.1\text{Total height}^Q - 0.564X \end{aligned} \quad (44)$$

$$X = \frac{\left( 1 - \sqrt[3]{\frac{h}{\text{Total height}}} \right)}{\left( 1 - \sqrt[3]{\frac{1.3}{\text{Total height}}} \right)} \text{ and } Q = 1 - \sqrt[3]{\frac{h}{\text{Total height}}}.$$

The elimination of the insignificant variables did not alter the excellent performance of the Kozak model 02, which had a correlation coefficient between the predicted and original data of 98% (Table 3). Among the considered stem taper

**Table 3** Comparison of the stem taper models for loblolly pine from east Louisiana

Measure	Parsimonious model		Kozak model 02	Bayesian model
	10 terms	30 terms		
Correlation coefficient (%)	96.1	96.4	98.1	96.3
Bias (%)				
Mean	-2	-0.1	-0.2	0.2
Variance	99.3	95.1	96.0	203.5
MAE (%)				
Mean	5.0	4.3	3.2	4.5
Variance	99.3	95.1	96.0	203.5
RMSE (%)				
Mean	8.2	7.7	3.8	8.1
Variance	99.7	95.2	96.0	209.9

models, the Kozak model 02 is the most suitable to describe the variation of diameter along the stem based on all the first-order moments, as the bias, *MAE* and *RMSE* were smaller than the parsimonious or Bayes models (Table 3). Nevertheless, when the second-order moment was considered, the Kozak 02 model exhibited larger values for all three measures (i.e., bias, *MAE*, and *RMSE*), but the difference was minute (Table 3). Both models, parsimonious and Kozak, exhibited no significant autocorrelation, as the Durbin–Watson test supplied a  $p$  value of  $> 0.05$ . Similar conclusion was reached for homoskedasticity, as Bartlett's test indicates that variance does not change with height ( $p$  value  $> 0.1$ ).

The parsimonious models with more than 20 terms in the Taylor expansion, the Bayesian model, and the Kozak 02 model are almost identical on the lower half of the stem (Fig. 3), the differences occurring on the upper portion, which inside the crown. Considering that the smallest volume is located inside the crown section of the stem, the differences between the models are operationally insignificant. The superior performance of the Kozak 02 model in respect to the first-order moments is not mirrored by the second-order moments, which place emphasis on the parsimony of the model rather than on the performances. Therefore, the less parsimonious model (i.e., Kozak 02) is more accurate but less precise than the more parsimonious model (Eq. 44).

## Discussion

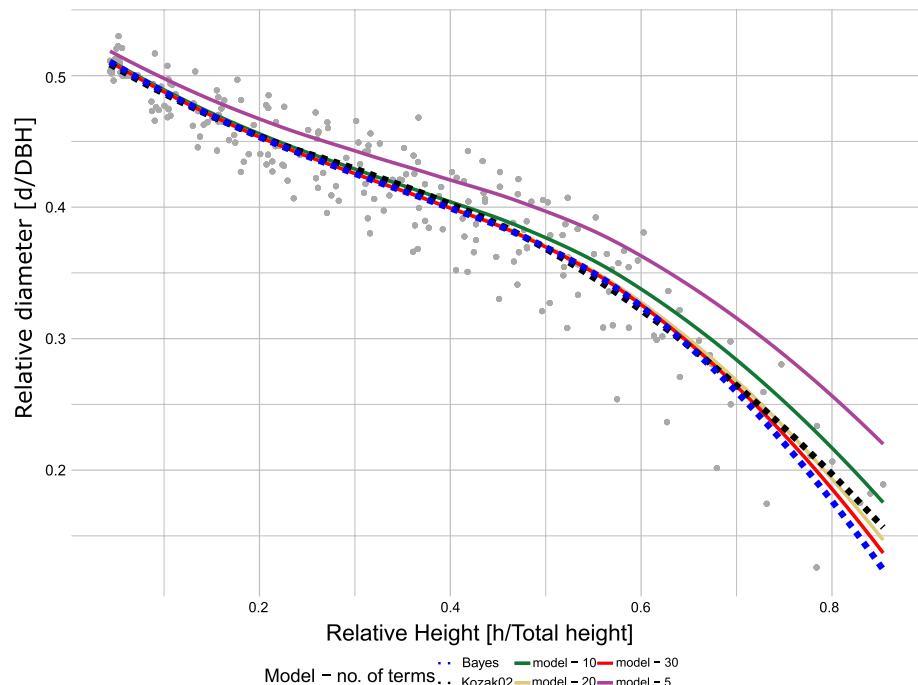
The parsimonious framework that we are proposing in this study completes the work of Strimbu et al (2017) and Amarioarei et al (2020) by presenting not only the first-order expectations but also the second order. Our models expand the findings of Neyman and Scott (1960) by providing an

approach for solving nonlinear models that is fast and simple. Our parsimonious approach is suitable for modeling complex nonlinear relationships, as a Taylor series expansion with 20 terms will provide numerically unbiased results. The successive application of transformation is particularly suitable for functions that do not include Taylor series, such as the hyperbolic arcsine. The improvement carried out by Nelder and Wedderburn (1972) on the work of Neyman and Scott (1960) simplified the distribution of residuals such that any distribution from the exponential family can be used. However, the generalization from normal to an exponential distribution proposed by Nelder and Wedderburn (1972) is based on fusing two distributions: of the residuals and of the predicted variable. The merging of residuals, which are a measure of the lack of knowledge, with a variable, precludes the applicability of generalized linear models to any function (Fox 2008). Consequently, many popular functions, such as trigonometric functions, cannot be used within the Nelder and Wedderburn (1972) framework because their expectation is difficult to compute. Our study restricts the distribution of residuals to normal distribution but allows inclusion of all the standard trigonometric functions in modeling. One of the main attractions of the proposed parsimonious approach is the ability to compound functions, such that exponential or power functions can be combined with trigonometric functions without biasing the results. Consequently, the linear equation (30) is arguably the simplest parsimonious method of solving nonlinear models. However, the requirement that the variance of the residuals must have small values in respect to the predicted variable indicates that the proposed

parsimonious modeling framework is highly dependent on data fitting.

The models for height of dominant and codominant Norway spruce and for the stem taper developed with the parsimonious framework presented in this study are accurate as the existing non-parsimonious models but more precise. The accuracy varies with the predicted variable, as the stem taper model is less accurate than the parsimonious model, whereas for the height the situation is reversed. Nevertheless, the differences are minute, which points toward the utility of the new framework. Even that the precision of the stem taper model seems to be superior to the parsimonious models, the variability of the non-parsimonious estimated *RMSE* suggests the opposite. A similar conclusion holds for height modeling, with the difference that the variability of the estimated assessment measures is significantly larger for the non-parsimonious models. Irrespective of the model, the non-parsimonious and the parsimonious models converge, but a large number of terms is needed in the Taylor series expansion, at least 20 (Figs. 2 and 3). The lack of parsimony is one aspect of the modeling process, another being the meaning of the variables predicting the variable of interest. From the meaning perspective, the parsimonious models include terms that can be interpreted [e.g., relative height (i.e.,  $h/\text{total height}$ ) or slenderness coefficient (i.e.,  $\text{total height}/\text{DBH}$ )], whereas the non-parsimonious models contain terms like  $\sqrt[10]{\frac{1 - \sqrt[3]{\text{relative height}}}{1 - \sqrt[3]{\frac{1.3}{\text{Total height}}}}}$ , which are not only difficult to infer or relate to tree processes, even that explanation for causality are argued to exist (LeMay 2018), but they are

**Fig. 3** The stem taper models for planted loblolly pine



usually present as a part of a series (i.e., the 30th root of a variable). Considering that the differences between the models are minute in terms of linear moments (i.e., accuracy), whereas the second-order moment is constantly smaller for parsimonious models, the model selection should be placed on parsimony. As the natural alternative to heuristic estimations of the nonlinear models, the Bayesian approach provided results inferior to the parsimonious models for all assessment metrics and moment order (i.e., linear or quadratic), which suggests that more research is needed in the Bayesian area to reach levels similar to the parametric approaches. One of the main challenges faced by the Bayesian approach is the reliance on computational resources, which for large datasets and complex nonlinear models could render solutions in unfeasible amount of time.

Our study provides a parsimonious method of solving complex nonlinear problems by enforcing the normality distribution of the residuals. The examples are used to show the performances of the proposed framework points toward superior results not only in terms of assessment metrics but also in terms of interpretability, particularly when the transformed linear model has a coefficient of determination larger than 0.95, recommended 0.99. The main limitation of the study rests in its normality assumption, as the formula are valid only when the residuals are normally distributed. Therefore, evidence of normality is required; we found that a  $p$  value  $> 0.25$  is recommended. Consequently, further studies are needed to expand the results to other distribution, such as the exponential family of distribution as defined by Nelder and Wedderburn (1972).

## Conclusions

The general linear regression framework developed by Galton (1877) was enhanced for more than one century but was not suitable for complex nonlinear models. The main issue faced by the nonlinear applications was the bias introduced by transforming the predicted variable. Our study builds on the work of Strimbu et al. (2017) and Amarioarei et al. (2020), and computes the bias corrections for the first and second order moments for some of the most popular transformations (i.e., power, trigonometric, and hyperbolic). Strimbu et al. (2017) proved that the estimated values are unbiased and complex nonlinear models can be obtained by compounding various transformations. Because the unbiased estimates for eight functions (i.e., sine, cosine, tangent, arcsine, arccosine, arctangent, hyperbolic sine, and hyperbolic tangent) contain a Taylor series expansion, we have shown that in most situations less than 10 terms are required for results without bias.

We have proved that the parsimonious framework can be applied successfully in forestry, by modeling the height

for dominant and codominant Norway Spruce from Romania and stem taper of loblolly pine from east Louisiana. We compare our results with the current models used for stem taper and height, both non-parsimonious. The parsimonious models contain two terms for height, compared with Giurgiu and Draghiciu (2004), which contains six terms, and three terms for stem taper, compared with nine of Kozak model 02 (2004). The parsimonious models have similar first-order expectation for the three-assessment metrics (i.e., bias, MAE, and RMSE) but smaller second-order moment, sometime close to one order of magnitude. Therefore, the parsimonious models, by relying on the normality assumption, gain in precision without a significant loss, if any, of accuracy. The Bayesian approach to the parsimonious nonlinear models exhibits similar first-order moments for the assessment metrics, but the second-order moments were significantly larger. The attractiveness of Bayesian solution is challenged by the size of the data and the complexity of the model, as it is computationally intensive. The subsequent studies would focus on generalizing the normal distribution for which the formulas are now available by computing the same estimates, but for the exponential family of distribution.

## Appendix

Using the same notation as above, the second moment  $\mathbb{E}[Y^2]$  is computed as following:

Power:  $g(y) = y^\alpha$  with  $\alpha > 0$  and  $y \in (0, \infty)$

$$\begin{aligned} \mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty y^2 e^{-\frac{1}{2}\left(\frac{y^\alpha - \xi}{\sigma}\right)^2} \alpha y^{\alpha-1} dy \stackrel{t = \frac{y^\alpha - \xi}{\sigma}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\frac{\xi}{\sigma}}^\infty (\xi + \sigma t)^{\frac{2}{\alpha}} e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \binom{\frac{2}{\alpha}}{k} \xi^{\frac{2}{\alpha}-k} \sigma^k \int_{-\frac{\xi}{\sigma}}^\infty t^k e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \binom{\frac{2}{\alpha}}{k} \xi^{\frac{2}{\alpha}-k} \sigma^k I(\alpha, k), \end{aligned}$$

where  $\alpha = -\frac{\xi}{\sigma}$ .

Sine:  $g(y) = \sin(y)$  with  $y \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$

$$\begin{aligned} \mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} y^2 e^{-\frac{1}{2}\left(\frac{\sin(y) - \xi}{\sigma}\right)^2} \cos(y) dy = \frac{1}{\sqrt{2\pi}\sigma} \int_{-1}^1 \arcsin^2(t) e^{-\frac{1}{2}\left(\frac{t - \xi}{\sigma}\right)^2} dt \\ &= \frac{1}{\sqrt{2\pi}\sigma} \sum_{n=0}^{\infty} \frac{4^n (n!)^2}{(2n+1)!(n+1)} \int_{-1}^1 t^{2n+2} e^{-\frac{1}{2}\left(\frac{t - \xi}{\sigma}\right)^2} dt \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{4^n (n!)^2}{(2n+1)!(n+1)} \sum_{k=0}^{2n+2} \binom{2n+2}{k} \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k) \end{aligned}$$

with  $\alpha = \frac{-1-\xi}{\sigma}$  and  $\beta = \frac{1-\xi}{\sigma}$ .

We used the following series expansion for  $\arcsin^2(t)$  for  $|t| < 1$

$$\arcsin^2(t) = \sum_{n=0}^{\infty} \frac{4^n (n! t^{n+1})^2}{(2n+1)!(n+1)!}.$$

Arcsine:  $g(y) = \arcsin(y)$  with  $y \in [-1, 1]$ .

Since

$$\cos(t) = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{(2n)!},$$

we have

$$\begin{aligned} \mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^2(t) e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \\ &= \frac{1}{2\sqrt{2\pi}\sigma} \left[ \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(2t) e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \right] \\ &= \frac{1}{2} \left[ \Phi(\beta) - \Phi(\alpha) - \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n 4^n}{(2n)!} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} t^{2n} e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \right] \\ &= \frac{1}{2} \left[ \Phi(\beta) - \Phi(\alpha) - \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n 4^n}{(2n)!} \sum_{k=0}^{2n} \binom{2n}{k} \xi^{2n-k} \sigma^k I(\alpha, \beta, k) \right], \end{aligned}$$

with  $\alpha = \frac{-\frac{\pi}{2} - \xi}{\sigma}$ ,  $\beta = \frac{\frac{\pi}{2} - \xi}{\sigma}$  and where the interchange between the summation and integration is a consequence of the application of the Bounded Convergence Theorem.

Cosine:  $g(y) = \cos(y)$  with  $y \in [0, \pi]$

$$\begin{aligned} \mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\pi} y e^{-\frac{1}{2}(\frac{\cos(y)-\xi}{\sigma})^2} |-\sin(y)| dy \\ &= -\frac{1}{\sqrt{2\pi}\sigma} \int_{-1}^1 \arccos^2(t) e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-1}^1 \left( \pi \arcsin(t) - \arcsin^2(t) - \frac{\pi^2}{4} \right) e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \\ &= \frac{1}{\sqrt{2\pi}\sigma} \left[ \pi \int_{-1}^1 \arcsin(t) e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt - \int_{-1}^1 \arcsin^2(t) e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \right] - \frac{\pi^2}{4} (\Phi(\beta) - \Phi(\alpha)) \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \left[ \frac{(2n-1)!!}{(2n)!!} \frac{\pi}{2n+1} \sum_{k=0}^{2n+1} \binom{2n+1}{k} \xi^{2n+1-k} \sigma^k I(\alpha, \beta, k) \right. \\ &\quad \left. - \frac{4^n (n!)^2}{(2n+1)!(n+1)} \sum_{k=0}^{2n+2} \binom{2n+2}{k} \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k) \right] - \frac{\pi^2}{4} (\Phi(\beta) - \Phi(\alpha)). \end{aligned}$$

Arccosine:  $g(y) = \arccos(y)$  with  $y \in [-1, 1]$

$$\begin{aligned} \mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\pi} \cos^2(t) e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \\ &= \frac{1}{2\sqrt{2\pi}\sigma} \left[ \int_0^{\pi} e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt + \int_0^{\pi} \cos(2t) e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \right] \\ &= \frac{1}{2} \left[ \Phi(\alpha) - \Phi(\beta) + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n 4^n}{(2n)!} \int_0^{\pi} t^{2n} e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \right] \\ &= \frac{1}{2} \left[ \Phi(\alpha) - \Phi(\beta) + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n 4^n}{(2n)!} \sum_{k=0}^{2n} \binom{2n}{k} \xi^{2n-k} \sigma^k I(\alpha, \beta, k) \right]. \end{aligned}$$

Tangent:  $g(y) = \tan(y)$  with  $y \in \left[0, \frac{\pi}{4}\right]$ .

Expanding  $\arctan(t)$  in Taylor series for  $|t| \leq 1$ , we have

$$\arctan(t) = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n+1}}{2n+1}.$$

Thus,

$$\begin{aligned} \mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^1 \arctan^2(t) e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^1 \left[ \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n+1}}{2n+1} \right]^2 e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \\ &= \frac{1}{\sqrt{2\pi}\sigma} \sum_{n,m=0}^{\infty} \frac{(-1)^{n+m}}{(2n+1)(2m+1)} \int_0^1 t^{2(n+m)+2} e^{-\frac{1}{2}(\frac{t-\xi}{\sigma})^2} dt \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n,m=0}^{\infty} \frac{(-1)^{n+m}}{(2n+1)(2m+1)} \sum_{k=0}^{2(n+m)+2} \binom{2(n+m)+2k}{k} \xi^{2(n+m)+2-k} \sigma^k I(\alpha, \beta, k) \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{m=0}^n \frac{(-1)^n}{(2m+1)(2n-2m+1)} \sum_{k=0}^{2n+2} \binom{2n+2k}{k} \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k) \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n (2H_{2n+1} - H_n)}{2n+1} \sum_{k=0}^{2n+2} \binom{2n+2}{k} \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k), \end{aligned}$$

where  $H_n = \sum_{k=0}^n \frac{1}{k}$  is the  $n$ th harmonic number.

Hyperbolic sine:  $g(y) = \sinh(y)$  with  $y \in [0, a]$ ,  $a < \infty$ .

Based on the expansion

$$\operatorname{arcsinh}(t) = \sum_{n=0}^{\infty} \frac{(-1)^n (2n)!}{2^{2n} (n!)^2} \frac{t^{2n+1}}{2n+1},$$

we obtain

$$\begin{aligned}
 E[Y^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^1 \operatorname{arcsinh}^2(t) e^{-\frac{1}{2}\left(\frac{t-\xi}{\sigma}\right)^2} dt \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \sum_{n=0}^{\infty} \frac{(-4)^n (n!)^2}{(n+1)(2n+1)!} \int_0^1 t^{2n+2} e^{-\frac{1}{2}\left(\frac{t-\xi}{\sigma}\right)^2} dt \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \sum_{n=0}^{\infty} \frac{(-4)^n (n!)^2}{(n+1)(2n+1)!} \sum_{k=0}^{2n+2} \binom{2n+2}{k} \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k)
 \end{aligned}$$

Hyperbolic arcsine:  $f(y) = \operatorname{arcsinh}(y)$  with  $y \in [0, \infty)$

$$\begin{aligned}
 \mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\infty} \sinh^2(t) e^{-\frac{1}{2}\left(\frac{t-\xi}{\sigma}\right)^2} dt \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\infty} \frac{\cosh(2t) - 1}{2} e^{-\frac{1}{2}\left(\frac{t-\xi}{\sigma}\right)^2} dt \\
 &= -\frac{1}{2} \left[ 1 - \Phi\left(-\frac{\xi}{\sigma}\right) \right] + \frac{1}{\sqrt{2\pi}\sigma} \int_{-\frac{\xi}{\sigma}}^{\infty} \cosh(2\xi + 2\sigma t) e^{-\frac{1}{2}\left(\frac{t-\xi}{\sigma}\right)^2} dt \\
 &= -\frac{1}{2} \left[ 1 - \Phi\left(-\frac{\xi}{\sigma}\right) \right] + \frac{e^{2\sigma^2}}{2} \left[ 2 \cosh(2\xi) - e^{2\xi} \Phi\left(-\frac{\xi}{\sigma} - 2\sigma\right) - e^{-2\xi} \Phi\left(-\frac{\xi}{\sigma} + 2\sigma\right) \right].
 \end{aligned}$$

Hyperbolic tangent:  $g(y) = \tanh(y)$  with  $y \in [0, \infty)$

$$\begin{aligned}
 \mathbb{E}[Y^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^1 \left[ \sum_{n=0}^{\infty} \frac{t^{2n+1}}{2n+1} \right]^2 e^{-\frac{1}{2}\left(\frac{t-\xi}{\sigma}\right)^2} dt \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \int_0^1 \sum_{n,m=0}^{\infty} \frac{t^{2n+2m+2}}{(2n+1)(2m+1)} e^{-\frac{1}{2}\left(\frac{t-\xi}{\sigma}\right)^2} dt \\
 &= \frac{1}{\sqrt{2\pi}} \sum_{n,m=0}^{\infty} \frac{1}{(2n+1)(2m+1)} \sum_{k=0}^{2n+2m+2} \binom{2n+2m+2}{2k} \xi^{2n+2m+2-k} \sigma^k I(\alpha, \beta, k) \\
 &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{2H_{2n+1} - H_n}{2n+1} \sum_{k=0}^{2n+2} \binom{2n+2k}{2k} \xi^{2n+2-k} \sigma^k I(\alpha, \beta, k)
 \end{aligned}$$

with  $H_n$  being the  $n$ th harmonic number and where we used the expansion

$$\operatorname{arctanh}(t) = \sum_{n=0}^{\infty} \frac{t^{2n+1}}{2n+1}.$$

**Acknowledgements** We would like to thank Dr. Valerie Lemay from University of British Columbia, Canada, and Dr. Dan Binkley from Northern Arizona University, USA, whose comments helped improving the manuscript.

**Author contributions** AA proved all the formulas and produced all the graphs and computations, with help from MP and BS. BS conceptualized the manuscript and wrote the manuscript with contributions from all co-authors. MP reviewed the computations and helped in designing the applications.

**Funding** This work was partially supported by the Romanian ANCSI project POC P-37-462 257, by the U.S. Department of Agriculture—McIntire Stennis project OREZ-1004, and by the U.S. Department of Agriculture, National Institute of Food and Agriculture, Grant Number 2019-67019-29462.

**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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Aledo JA, Gámez JA, Molina D (2016) Using metaheuristic algorithms for parameter estimation in generalized Mallows models. *Appl Soft Comput* 38:308–320. <https://doi.org/10.1016/j.asoc.2015.09.050>

Amarioarei A, Paun M, Strimbu B (2020) Development of nonlinear parsimonious forest models using efficient expansion of the Taylor series: applications to site productivity and taper. *Forests* 11:458. <https://doi.org/10.3390/f11040458>

Ansley CF, Kohn R, Shively TS (1992) Computing p-values for the generalized Durbin–Watson and other invariant test statistics. *J Econom* 54:277–300. [https://doi.org/10.1016/0304-4076\(92\)90109-5](https://doi.org/10.1016/0304-4076(92)90109-5)

Bartlett MS (1937) Properties of sufficiency and statistical tests. *Proc R Soc Lond Ser Math Phys Sci* 160:268–282

Bettinger P, Graetz D, Boston K et al (2002) Eight heuristic planning techniques applied to three increasingly difficult wildlife planning problems. *Silva Fenn* 36:561–584

Bilskie MV, Hagen SC (2013) Topographic accuracy assessment of bare earth lidar-derived unstructured meshes. *Adv Water Resour* 52:165–177. <https://doi.org/10.1016/j.advwatres.2012.09.003>

Brockwell PJ, Davis RA (1996) An introduction to time series and forecasting. Springer, New York

Cao QV, Burkhart HE, Max TA (1980) Evaluation of 2 methods for cubic-volume prediction of loblolly-pine to any merchantable limit. *For Sci* 26:71–80

Cattell RB (1966) Scree test for number of factors. *Multivar Behav Res* 1:245–276. [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10)

Cochran WG (1938) Some difficulties in the statistical analysis of replicated experiments. *Emp J Exp Agric* 15:157–175

Cotes R (1722) Harmonia mensurarum. Cantabrigienfes Socius, Cantabrigie UK

Duursma RA, Robinson AP (2003) Bias in the mean tree model as a consequence of Jensen's inequality. *For Ecol Manag* 186:373–380. [https://doi.org/10.1016/S0378-1127\(03\)00307-4](https://doi.org/10.1016/S0378-1127(03)00307-4)

Fang R, Strimbu B (2017) Stem measurements and taper modeling using photogrammetric point clouds. *Remote Sens* 9:21

Fang Z, Borders BE, Bailey RL (2000) Compatible volume-taper models for loblolly and slash pine based on a system with segmented-stem form factors. *For Sci* 46:1–12. <https://doi.org/10.1093/forests/science/46.1.1>

Finney DJ (1941) On the distribution of a variate whose logarithm is normally distributed. *J R Stat Soc Ser B* 7:155–161

Fox J (2008) Applied regression analysis and generalized linear models, 2nd edn. SAGE Publications, Thousand Oaks, CA

Galton F (1877) Typical laws of heredity. *Nature* 15:492–495

Gelman A, Carlin JB, Stern HS, Rubin DB (2003) Bayesian data analysis, 2nd edn. Chapman and Hall, Boca Raton, FL

Gentleman R, Ihaka R (2014) R. University of Auckland, Auckland, New Zealand

Giurgiu V, Draghiciu D (2004) Modele matematico-auxologice și tabele de producție pentru arborete. Ceres, Bucharest

Golivets M, Woodall CW, Wallin KF (2019) Functional form and interactions of the drivers of understory non-native plant invasions in northern US forests. *J Appl Ecol* 56:2596–2608. <https://doi.org/10.1111/1365-2664.13504>

Grimmett GD, Stirzaker DR (2002) Probability and random processes. Oxford University Press, New York, NY

Hoos H, Stutzle T (2005) Stochastic local search. Morgan Kaufmann Publishers, New York

Chen J, Kemna A, Hubbard SS (2008) A comparison between Gauss–Newton and Markov-chain Monte Carlo-based methods for inverting spectral induced-polarization data for Cole–Cole parameters. *Geophysics* 73:F247–F259. <https://doi.org/10.1190/1.2976115>

Kansanen K, Vauhkonen J, Lähivaara T et al (2019) Estimating forest stand density and structure using Bayesian individual tree detection, stochastic geometry, and distribution matching. *ISPRS J Photogramm Remote Sens* 152:66–78. <https://doi.org/10.1016/j.isprsjprs.2019.04.007>

Korzukhin MD, TerMikaelian MT, Wagner RG (1996) Process versus empirical models: which approach for forest ecosystem management? *Can J For Res-Rev Can Rech For* 26:879–887

Kozak A (2004) My last words on taper equations. *For Chron* 80:507–515. <https://doi.org/10.5558/tfc80507-4>

LeMay V (2018) Personal communication on the presentation “modeling in the age of big-data and AI: the loss of beauty” by Strimbu BM

Lenhart JD, Hackett TL, Laman CJ et al (1987) Tree content and taper functions for loblolly and slash pine trees planted on non-old-fields in East Texas. *South J Appl For* 11:147–151

Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. *Q Appl Math* 2:164–168. <https://doi.org/10.1090/qam/10666>

LeVeque RJ (2007) Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems. SIAM, Philadelphia, PA

Lindstrom MJ, Bates DM (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics* 46:673–687. <https://doi.org/10.2307/2532087>

Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 11:431–441. <https://doi.org/10.1137/0111030>

Max TA, Burkhart HE (1976) Segmented polynomial regression applied to taper equations. *For Sci* 22:283–289

McClure JP, Czaplewski RL (2011) Compatible taper equation for loblolly pine. *Can J For Res.* <https://doi.org/10.1139/x86-225>

Montealegre A, Lamelas M, Riva J (2015) Interpolation routines assessment in ALS-derived digital elevation models for forestry applications. *Remote Sens* 7:8631

Nelder JA, Wedderburn RWM (1972) Generalized linear models. *J R Stat Soc Ser Gen* 135:370–384. <https://doi.org/10.2307/2344614>

Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) Applied linear statistical models. WCB McGraw-Hill, Boston

Newton I (1687) Philosophiae Naturalis Principia Mathematica. Jussu Societatis Regiae, London UK

Neyman J, Scott EL (1960) Correction for bias introduced by a transformation of variables. *Ann Math Stat.* <https://doi.org/10.1214/aoms/1177705791>

Nicoletti MF, de Pádua Chaves e Carvalho S, do Amaral Machado S et al (2020) Bivariate and generalized models for taper stem representation and assortments production of loblolly pine (*Pinus taeda* L.). *J Environ Manage* 270:110865. <https://doi.org/10.1016/j.jenvman.2020.110865>

Özsoy VS, Ünsal MG, Örkü HH (2020) Use of the heuristic optimization in the parameter estimation of generalized gamma distribution: comparison of GA, DE, PSO and SA methods. *Comput Stat* 35:1895–1925. <https://doi.org/10.1007/s00180-020-00966-4>

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554. <https://doi.org/10.1093/biomet/58.3.545>

Paun M, Gunaim N, Strimbu BM (2020) Impact of algorithm selection on modeling ozone pollution: a perspective on Box and Tiao (1975). *Forests* 11:1311. <https://doi.org/10.3390/f11121311>

Popescu-Zeletin I (1957) Tabele dendrometrice. Editura Agrosilvica de Stat, Bucharest

Prieto-Escobar N, Saldarriaga-Aristizábal PA, Chaparro-Muñoz V et al (2018) Heuristic parameter estimation for a continuous fermentation bioprocess. *Rev Fac Ing Univ Antioquia.* <https://doi.org/10.17533/udea.redin.n88a04>

Pujol J (2007) The solution of nonlinear inverse problems and the Levenberg–Marquardt method. *Geophysics* 72:W1–W16. <https://doi.org/10.1190/1.2732552>

Schumacher FX, Hall FDS (1933) Logarithmic expression of timber-tree volume. *J Agric Res* 47:719–734

Seppelt R, Richter O (2005) “It was an artefact not the result”: a note on systems dynamic model development tools. *Environ Model Softw* 20:1543–1548

Shanks ME, Gambill R (1973) Calculus. Holt, Rinehart and Winston, Inc., New York

Stan Development Team (2016a) Stan modeling language users guide and reference manual. Version 2.15.0.

Stan Development Team (2016b) Rstan: the R interface to Stan

Stängle SM, Sauter UH, Dormann CF (2017) Comparison of models for estimating bark thickness of *Picea abies* in southwest Germany: the role of tree, stand, and environmental factors. *Ann For Sci* 74:16. <https://doi.org/10.1007/s13595-016-0601-2>

Stow CA, Reckhow KH, Qian SS (2006) A Bayesian approach to retransformation bias in transformed regression. *Ecology* 87:1472–1477. [https://doi.org/10.1890/0012-9658\(2006\)87\[1472:Abatr\]2.0.Co;2](https://doi.org/10.1890/0012-9658(2006)87[1472:Abatr]2.0.Co;2)

Strimbu BM, Amarioarei A, Paun M (2017) A parsimonious approach for modeling uncertainty within complex nonlinear relationships. *Ecosphere* 8:e01945

Tabachnick BG, Fidell LS (2001) Using multivariate statistics. Allyn and Bacon, Needham Heights

Talbi E-G (2009) Metaheuristics: from design to implementation. Wiley, Hoboken, NJ

Taylor B (1715) Methodus Incrementorum Directa et Inversa. Gulienini Innys, London, UK

Tudoran GM, Zotta M (2020) Adapting the planning and management of Norway spruce forests in mountain areas of Romania to environmental conditions including climate change. *Sci Total Environ* 698:133761. <https://doi.org/10.1016/j.scitotenv.2019.133761>

Valentine HT, Gregoire TG (2001) A switching model of bole taper. *Can J For Res* 31:1400–1409. <https://doi.org/10.1139/x01-061>

van Oijen M (2017) Bayesian methods for quantifying and reducing uncertainty and error in forest models. *Curr For Rep* 3:269–280. <https://doi.org/10.1007/s40725-017-0069-9>

von Leibniz GWF (1920) The early mathematical manuscripts of Leibniz. The Open Court Publishing company, Chicago IL, USA

Warton DI, Hui FKC (2011) The arcsine is asinine: the analysis of proportions in ecology. *Ecology* 92:3–10. <https://doi.org/10.1890/10-0340.1>

Williams CB (1937) The use of logarithms in the interpretation of certain entomological problems. *Ann Appl Biol* 24:404–414. <https://doi.org/10.1111/j.1744-7348.1937.tb05042.x>

Yuan Y (2015) Recent advances in trust region algorithms. *Math Program* 151:249–281. <https://doi.org/10.1007/s10107-015-0893-2>

Yuan YX (2011) Recent advances in numerical methods for nonlinear equations and nonlinear least squares. *Numer Algebra Control Optim* 1:15–34

Zhu M, Ghodsi A (2006) Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Comput Stat Data Anal* 51:918–930. <https://doi.org/10.1016/j.csda.2005.09.010>

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