Scientific
Research
Publishing

Growth Curves for Diameter and Height Using Mixed Models: An Application in Eucalyptus Seedling

Giovana Fumes[1]*, Clarice Garcia Borges Demétrio[1], Cristian Villegas[1], José Eduardo Corrente[2], Juliane Fumes Bazzo[3]

[1]Department of Exact Sciences, ESALQ, University of São Paulo, Piracicaba, Brazil
[2]Department of Biostatistics, University of São Paulo State, Botucatu, Brazil
[3]Department of Natural Resources-Forest Sciences, University of São Paulo State, Botucatu, Brazil
Email: *gifumesbtu@gmail.com

Abstract

The growth of a tree or a forest settlement is of great value to a forest enterprise, because many decisions are directly dependent of this information, for instance, determining the optimal cutting age. This study aims to apply a new class of models to fit growth curves for diameter and height of *Eucalyptus grandis X Eucalyptus urophylla* seedling data. Data were collected from a trial conducted in a green house at the Natural Resources Department at School of Agriculture, Botucatu, São Paulo, Brazil. The experiment's design was completely randomized with eight treatments and four replications. In this trial, the growth variables referring to the height and the diameter were evaluated, being measured five and four times, respectively. The methodology was carried in a mixed longitudinal model using a new approach based on Box-Cox Normal (BCN) distribution, and comparisons with this model were made assuming normality of the data. The results revealed that the BCN mixed model provided similar results to the standard model in order to estimate growth curves; however, the BCN model was the best result according to Akaike criterion, considering the slight asymmetry in the data set. This approach is of great interest in case of outliers and robust procedures for parameter estimation.

1. Introduction

The increase in wood consumption and its derivatives highlight the need to gen-

erate new seedlings production technologies with a standard of appropriate quality to establish more productive forests.

In the seedlings selection for planting, the criteria are based on characteristics which generally do not determine the real quality, because the seeds vary according to the species, ecological sites, cultivation, transportation, distribution and planting. So, there are several reasons for the use of tests in the standard definition of seedlings quality, being able to add some values that are often required by the market (Gomes et al., 2002).

The survival, the establishment, the cultivation frequency and the early forest growth are necessary ratings for the forest enterprise success, which is directly related to the seedling quality at planting (Gomes, 2001).

In this context, the height/diameter relation is one of the features used to evaluate the forest seedling quality, because it reflects the reserve accumulation, a greater resistance and better fixation in the soil (Carneiro, 1995). Furthermore, the average height and the base diameter are two parameters, which show the high plant quality considered relevant to forestry industry companies (Gomes et al., 1996).

A set of mathematical models for growth in height and in diameter was proposed using linear and nonlinear models by many authors, including Schumacher (1939), Bertalanffy (1951), Richards (1959), Prodan (1968), Machado (1979), Silva (1986), which are known as Brody, von Bertalanffy, logistic and Gompertz models. These models have been widely and successfully used in many works, such as in Machado et al. (1997), Dias et al. (2005), Oliveira et al. (2008), Teo et al. (2011), Carvalho et al. (2014). Such models assume normal distribution for the dataset but, generally this assumption is not verified.

Ferrari & Fumes (2017) proposed a new class of distributions called Box-Cox symmetric class, containing distributions that allow it to model asymmetry and that can also include outliers. Special cases include log-symmetric distribution (Vanegas & Paula, 2015), (for instance, log-normal distribution) and the truncated symmetric distributions with support on the positive real line, such as the Box-Cox t (Rigby & Stasinopoulos, 2006), the Box-Cox Normal (Cole & Green, 1992; Stasinopoulos et al., 2008) and the Box-Cox power exponential distributions (Voudouris et al., 2012), among others.

Therefore, the aim of this work is to apply models taken into account this new distribution class to fit growth curves for height and diameter data from a forestry trial of the *Eucalyptus grandis X Eucalyptus urophylla* hybrid seedlings and compare them with models that assume normality of the data.

2. Materials and Methods

2.1. The Data Set

This work arised from the data collected from a trial carried out in a green house of the Department of Natural Resources - Forestry Sciences, School of Agriculture, UNESP, Botucatu, São Paulo, Brazil (Bazzo, 2009). The experiment eva-

luated the effects of different substrates on the *Eucalyptus grandis X Eucalyptus urophylla* hybrid seedling development. Different proportions of sewage sludge and carbonized rice husk, which composed the substrates, were compared to a standard substrate used in the green house. For the substrate composition, the sewage used sludge by carbonized rice husk proportions were: 100:0, 80:20, 60:40, 50:50, 40:60, 20:80 and 0:100. For this article, the different proportions composed the treatments from one to seven, in the order of proportions as shown above, and the eighth treatment was the standard substrate. The experimental design was a completely randomized design with eight treatments and four replications, each replication being origined by the growth average of 24 plantings. The growth variables measured were the height in five periods (after 90, 105, 120, 135 and 150 days) and the diameter in four periods (after 105, 120, 135 and 150 days), both measurement in centimeters.

Figure 1 revealed a linear growth trend for height (Figure 1(a)) and nonlinear growth trend for diameter (Figure 1(b)).

Figure 2 shows the mean profile analysis. We observed that the eighth
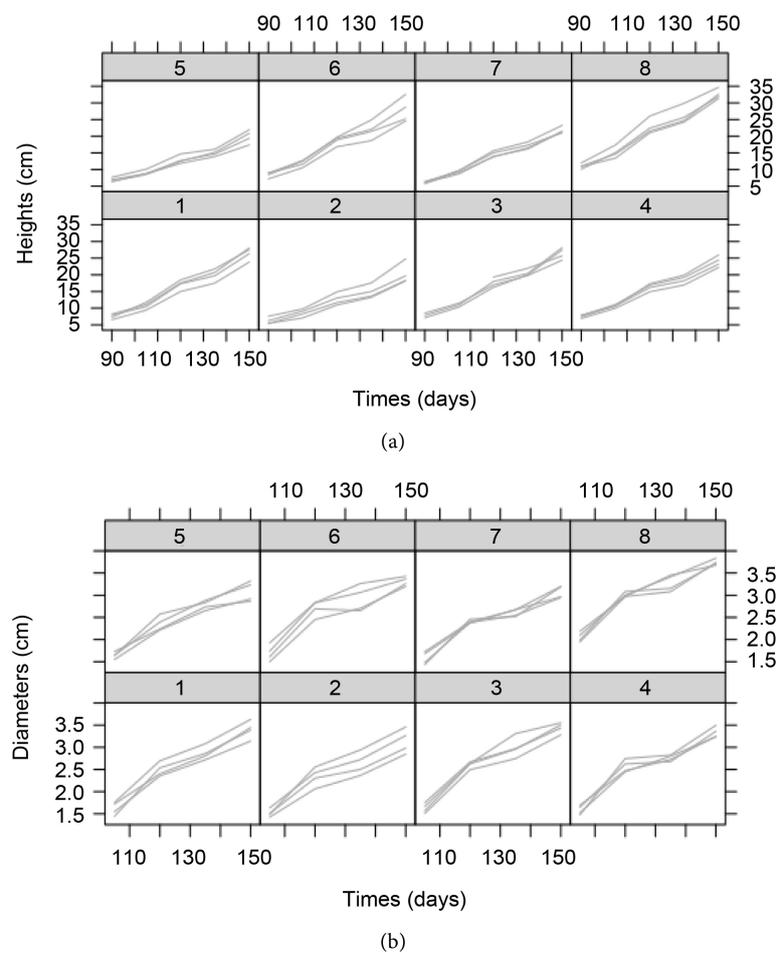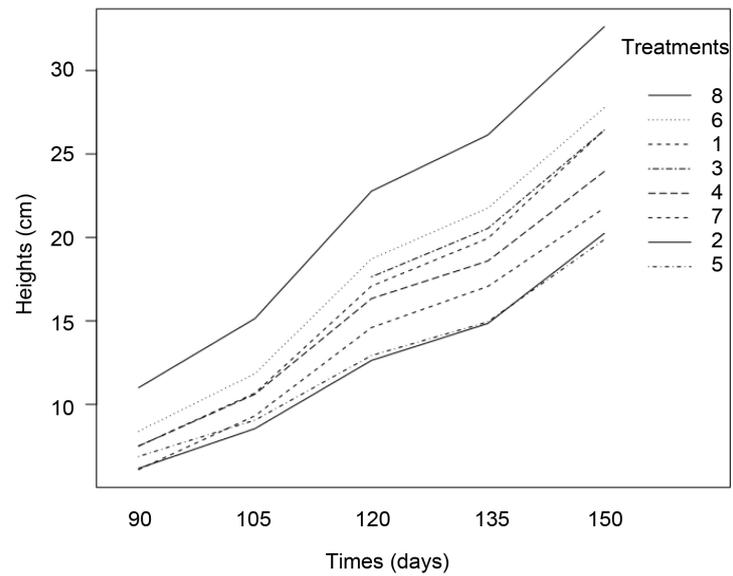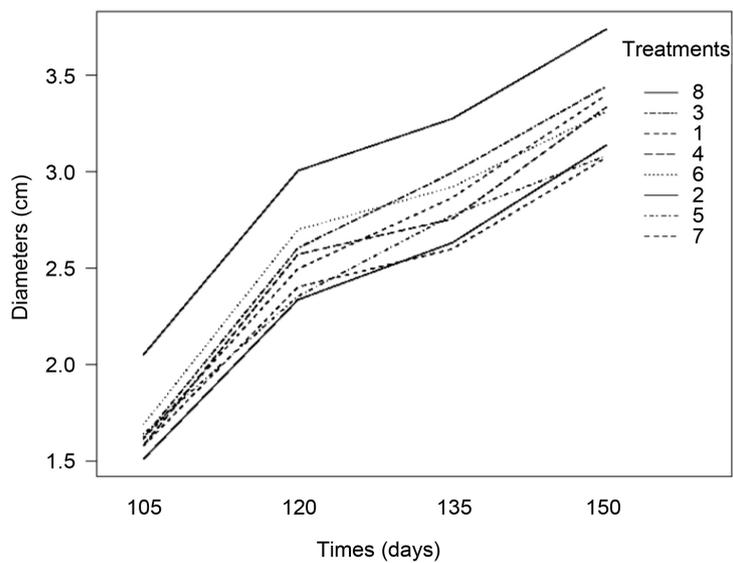

(a)

(b)

Figure 1. Profile analysis of each treatment. The sequence of the treatment begins from bottom left (Treatment 1) to top right (Treatment 8). (a) Growth in height. (b) Growth in diameter.

Figure 2. Mean profile analysis. (a) Growth in height. (b) Growth in diameter.

treatment, which was the standard compost, presented a better growth performance.

The histograms for the height data by time in **Figure 3** showed symmetrical (**Figure 3(c)**, **Figure 3(d)**) and slightly asymmetrical (**Figure 3(a)**, **Figure 3(b)**, **Figure 3(e)**) shapes that were different in each time. Plots for the diameter data were omitted because they presented the same feature.

2.2. (Non) Linear Mixed Model Approach

In this section, it was described the (non) linear mixed model approach for height and diameter modelling that was used to compare it with a new approach
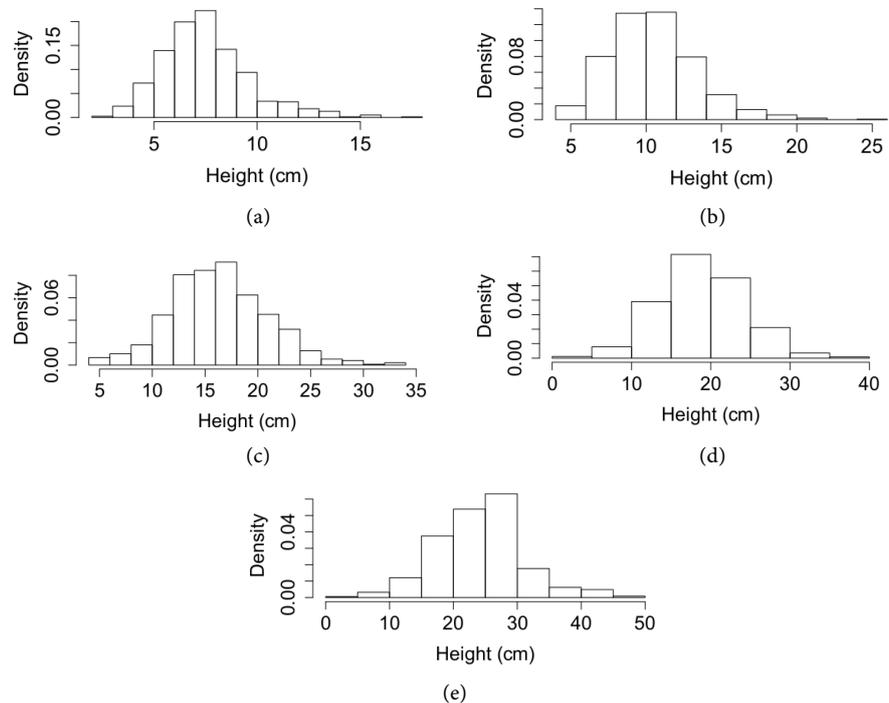
Figure 3. Histograms by time for the height variable. (a) Time 1. (b) Time 2. (c) Time 3. (d) Time 4. (e) Time 5.

based in the Box-Cox symmetric class (Ferrari & Fumes, 2017).

Having problems involving growth data, the structure of linear mixed model is ordinarily used to model longitudinal data (Melesse & Zewotir, 2017; Soares et al., 2017; Pinheiro & Bates, 2000). In this case, the height and the diameter data were assumed to be normally distributed and the independence assumption was violated in longitudinal data, hence random effects were required.

So, for the height data, which presented a linear trend in descriptive analysis (Section 2.1), the model was defined by:

$$Y_{ijk} = \beta_{i1} + \beta_{i2}t_k + \epsilon_{ij} + \varepsilon_{ijk},$$

(1)

where Y_{ijk} was the height related to the kth time, the jth replication and the ith treatment, β_{i1} was the mean of the ith treatment, β_{i2} was the slope parameter of the ith treatment, t_k was the time (continuous), $\epsilon_{ij} \sim N\left(0,\sigma_p^2\right), iid,$ was the plot random effect and $\varepsilon_{ijk} \sim N\left(0,\sigma^2\right), iid,$ was the error of the model, for $i = 1, \cdots, 8$, $j = 1, \cdots, 4$ and $k = 1, \cdots, 5$.

For the diameter data, which presented a nonlinear trend in the descriptive analysis (Section 2.1), the Gompertz mixed model was proposed with the structure:

$$Y_{ijk} = \left(\beta_1 + \epsilon_{ij}\right)\exp\left(-\beta_2\beta_2^{t_k}\right) + \varepsilon_{ijk},$$

(2)

where Y_{ijk} was the diameter related to the kth time, the jth replication and the ith treatment, β_1 was the parameter representing the asymptote, $\epsilon_{ij} \sim N\left(0,\sigma_p^2\right), iid;$ β_2 was the parameter related to the value of the function at

$t_k = 0$, β_3 was the parameter related to the scale of the t variable and t_k was the time (continuous), and $\varepsilon_{ijk} \sim N(0,\sigma^2), iid,$ was the error of the model, for $i = 1,\cdots,8$, $j = 1,\cdots,4$ and $k = 1,\cdots,4$.

The components of variance parameters of the linear mixed model were estimated by the restricted maximum likelihood (REML). This methodology is a particular approach form of maximum likelihood estimation which does not calculate estimates on a maximum likelihood fit of all the information, but instead, uses a likelihood function calculated from a transformed set of data. REML can produce less unbiased estimates of variance and covariance parameters (Pinheiro & Bates, 2000, Chapter 2). The penalized nonlinear least squares (PNLS) algorithm was used for nonlinear model (Pinheiro & Bates, 2000, Chapter 7). All the procedures were in R (R Core Team, 2015), version 3.3.1, using *lme* 4 for linear and *nlme* for nonlinear models routines (Pinheiro & Bates, 2000). For the goodness of fit, the standardized residual was used for linear and nonlinear models (Pinheiro & Bates, 2000).

2.3. The BCN Model Approach

Ferrari & Fumes (2017) proposed a new class of distributions called Box-Cox symmetric class, which includes the Box-Cox Normal distribution, that can be fitted to data with symmetric and asymmetric shapes. Then, an alternative approach, using the same structure of linear mixed model for the height data, was:

$$Y_{ijk} \mid \epsilon_{ij} \sim BCN(\mu_{ijk}, \sigma, \nu), \text{ where } \mu_{ijk} = \beta_{i1} + \beta_{i2} t_k + \epsilon_{ij}, \tag{3}$$

where Y_{ijk} was the height related to the kth time, the jth replication and the ith treatment with Box-Cox Normal (BCN) distribution, where μ_{ijk} was associated with: the median of the ith treatment (β_{i1}), β_{i2} was the slope parameter of the ith treatment, t_k was the time (continuous) and $\epsilon_{ij} \sim N(0,\sigma_p^2), iid,$ was the plot random effect, for $i = 1,\cdots,8$, $j = 1,\cdots,4$ and $k = 1,\cdots,5$; besides, σ was the parameter related to the coefficient of variation based on the median and ν was the parameter associated to the asymmetry, with $\mu_{ijk} > 0$, $\sigma > 0$ and $\nu \in \mathbb{R}$.

For the diameter data, another alternative approach using BCN model was:

$$Y_{ijk} \mid \epsilon_{ij} \sim BCN(\mu_{ijk}, \sigma, \nu), \text{ where } \mu_{ijk} = \beta_i + f(t_k) + \epsilon_{ij}, \tag{4}$$

where Y_{ijk} was the diameter related to the kth time, the jth replication and the ith treatment with a Box-Cox Normal (BCN) distribution, where μ_{ijk} was associated with: the median of the ith treatment (β_i), $f(t_k)$ was the smooth function (in this case, p-spline smoother was used, see Stasinopoulos et al. (2008) for details) which models the time effects, $\epsilon_{ij} \sim N(0,\sigma_p^2), iid,$ was the plot random effect; σ was the parameter related to the coefficient of variation based on the median and ν was the parameter associated to the asymmetry, for $i = 1,\cdots,8$, $j = 1,\cdots,4$ and $k = 1,\cdots,4$; with $\mu_{ijk} > 0$, $\sigma > 0$ and $\nu \in \mathbb{R}$.

The marginal maximum likelihood was used for BCN models. In this context, the main distribution was been marginalized (Rigby et al., 2014). All the proce-

dures were performed in R (R Core Team, 2015), version 3.3.1, using *gamlss* routine for linear as well as semiparametric BCN models (Rigby et al., 2014). For the goodness of fit, the quantile residual plot was calculated for BCN models (Rigby et al., 2014).

In these approaches, the random effect ϵ_{ij} represented the between plot variability in the experiment, while the within plot variability was modelled with the variance of the residual error ε_{ijk} in the standard models, and in the BCN, it was modelled through the parameter σ. The Akaike criteria were calculated to compare the different approaches.

3. Results

The BCN and the normal distributions were plotted on top of the raw data in each time. For the height data, the BCN distribution performed better than the normal distribution (Figure 4). Plots for the diameter data were omitted because they presented the same feature.

In Table 1, there were the mean and the median estimated of growth for height by treatment for linear models (1) and (2). The estimated means were smaller than the estimated median. This result was expected due to the slight asymmetry presented in the data. In the model (4), it was possible to observe the estimated intercept for each treatment, and in this case, the smooth function described the growth curve for diameter.

For nonlinear mixed model (model (3)), the estimated parameters for Gompertz were: $\hat{\beta}_1 = 3.51$, $\hat{\beta}_2 = 1.55$ and $\hat{\beta}_3 = 0.47$ for fixed effects and
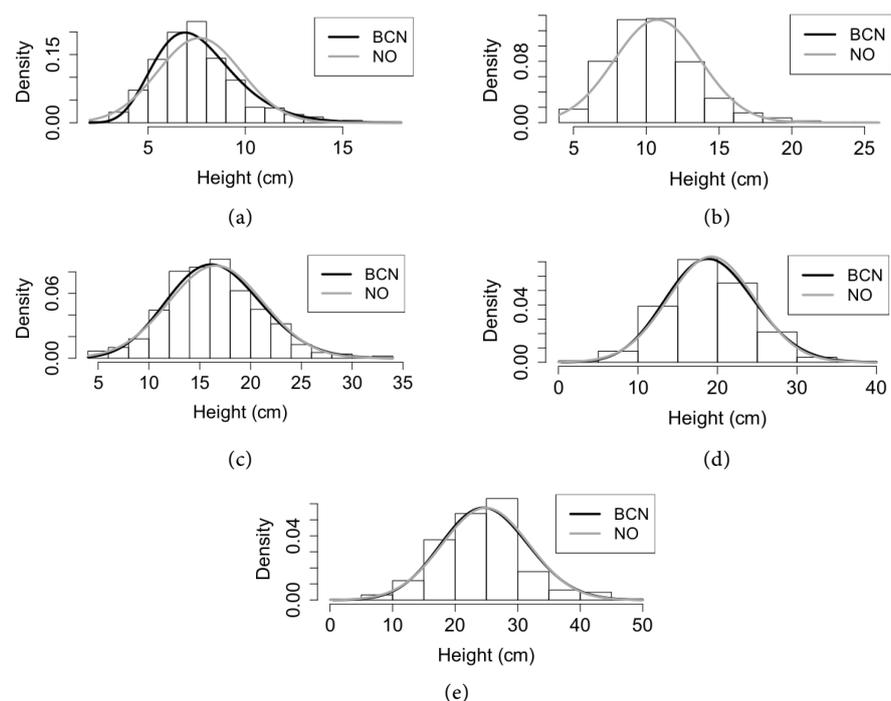

Figure 4. Raw data and densities for normal and BCN distributions for the height variable. (a) Time 1. (b) Time 2. (c) Time 3. (d) Time 4. (e) Time 5.

Table 1. Estimated parameters for fixed effects of linear model and linear and nonlinear BCN mixed models.

Treatments (i)	Linear mixed model (1)		Linear BCN mixed model (2)		Nonlinear BCN mixed model (4)
	β_{i1}	β_{i2}	β_{i1}	β_{i2}	β_i
1	2.14	4.17	2.76	4.46	1.14
2	2.16	3.44	2.84	3.15	0.99
3	2.61	4.70	2.72	3.61	1.20
4	3.14	4.09	3.38	3.98	1.12
5	3.18	3.19	3.69	2.98	1.08
6	3.10	4.87	3.51	4.65	1.22
7	2.04	3.91	2.14	3.87	1.03
8	5.26	5.42	5.42	5.33	1.58
|AIC|	584.15		549.98		52.08

$|AIC| = 111.51$. It means that the estimate of the parameter associated to asymptotic growth for this observed period was around 3.5 cm, with an estimated growth efficiency of 0.47 cm in each period. From AIC criteria, the BCN model had a better performance than the standard ones, even though we could observe that the parameter estimate were quite close as well as the AIC (models (1) and (2)). This occurred probably due to the presence of slight asymmetry already commented above, but still indicating the BCN can be considered superior than the other.

For diameter, we can observe that the nonlinear BCN mixed model (model (4)) presented lower AIC than the Gompertz model (model (3)). It is important to notice that in the BCN model, a smooth function was taking into account to obtain a better fit. In this case, the smooth function can lead to a better fit but it had a cost of lack of parsimony while Gompertz model uses only three parameters. We guess this sort of modelling can be better investigated in order to get a better fit with few parameters.

Figure 5 shows the plots of standardized residuals for the fitted models to the height and diameter data. We observed that the linear and nonlinear models had a good fit for the height and diameter variables, respectively. For the BCN models (Figure 6), it was explored the quantile residual as a result of implementation of the *gamlss* routines, whereby we also observed a good fit.

Figure 7 presents the plots of estimated lines of growth for the height data using normal and BCN models and shows that both models were adequate to model the data, although they have different estimation approaches. The linear model is based in the mean and standard deviation around the mean (Figure 7(a)). On the other hand, the BCN model describes the median and a coefficient of the variation around the median (Figure 7(b)). In this context, we observed that there was less variability in the BCN model than in the normal model, as showed by the growth lines. These lines for the BCN model were more concentrated due to the robustness in the estimation process of the parameters.
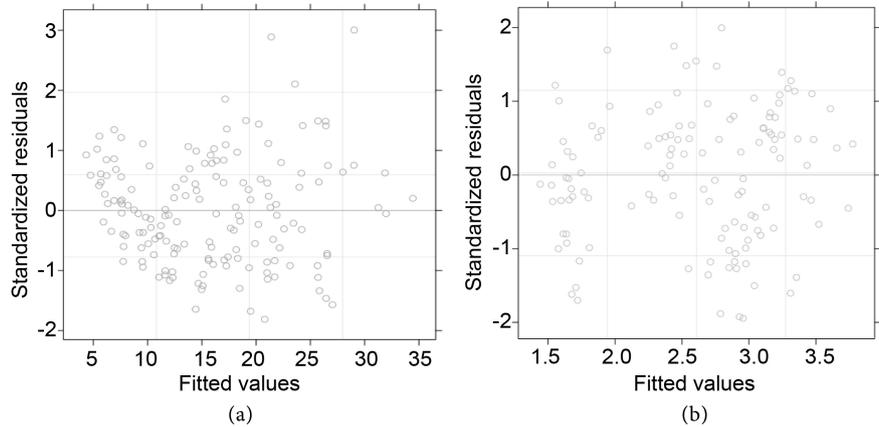
Figure 5. Plots of standardized residuals versus fitted values for the height (a) and diameter (b) data. (a) Linear mixed-effect model (1). (b) Nonlinear mixed-effect model (2).
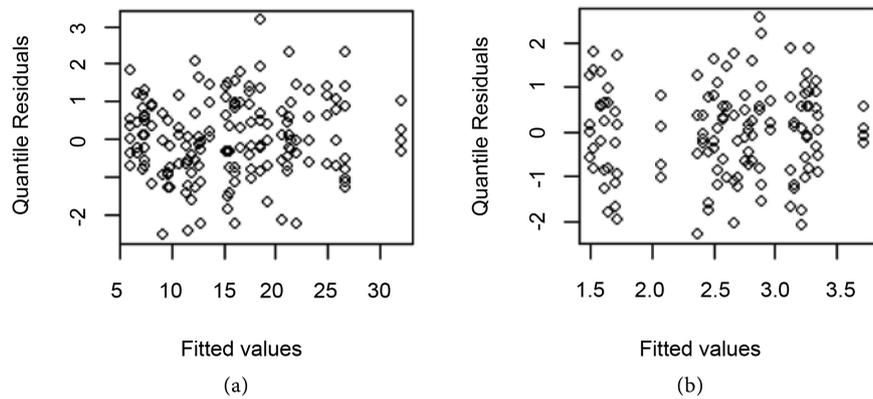

Figure 6. Plots of quantile residuals versus fitted values for the height (a) and diameter (b) data. (a) Linear BCN mixed-effect model (3). (b) Nonlinear BCN mixed-effect model (4).
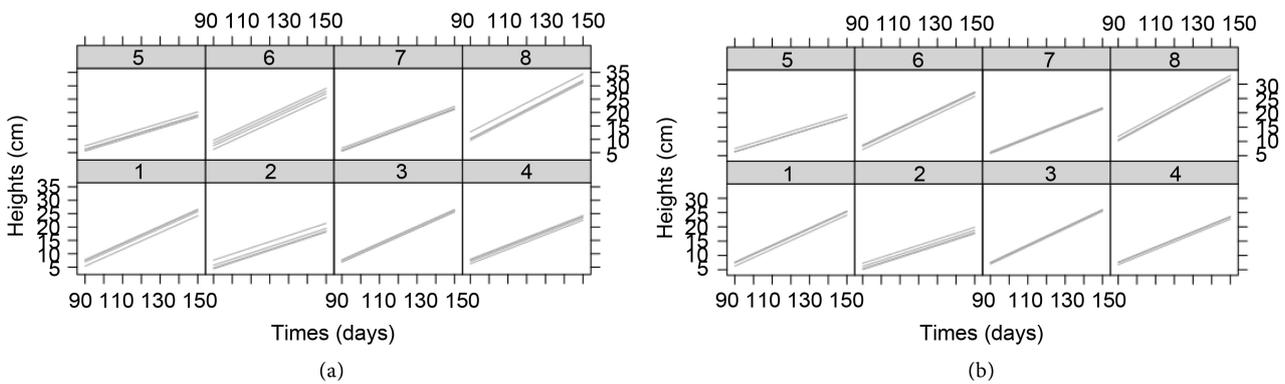

Figure 7. Fitted growth lines of height for each treatment. The sequence of the treatment begins from bottom left (Treatment 1) to top right (Treatment 8). (a) Linear mixed-effect model (1). (b) Linear BCN mixed-effect model (3).

Figure 8 shows the plots of the estimated curves for growth in diameter. The Gompertz mixed model (**Figure 8(a)**) reasonably described the data behaviour, but the Gompertz model has a specific mode and it did not show

completely the behaviour of the data. However, the BCN model showed a more realistic trend of growth (**Figure 8(b)**). The main reason is that the smooth function in the BCN model tends to model the data performance. Analogously to the linear model, there was less variability in the BCN model than in the nonlinear model because of the robust estimation.

Furthermore, for all the models, comparisons among the eight treatments were done. The eighth treatment, which was the standard compound, was statistically different from the others with a better growth performance (**Figure 9** and **Figure 10**). This result agrees with the descriptive analysis (**Figure 2**). Although the other treatments did not show such growth as the standard, they yielded satisfactory results. This is very important, because the treatments based on compounds by sewage sludge mixture are sanitarily safe for agriculture and a promising alternative to recycle municipal waste.

4. Conclusion

This paper presented a new approach based on the Box-Cox Normal distribution

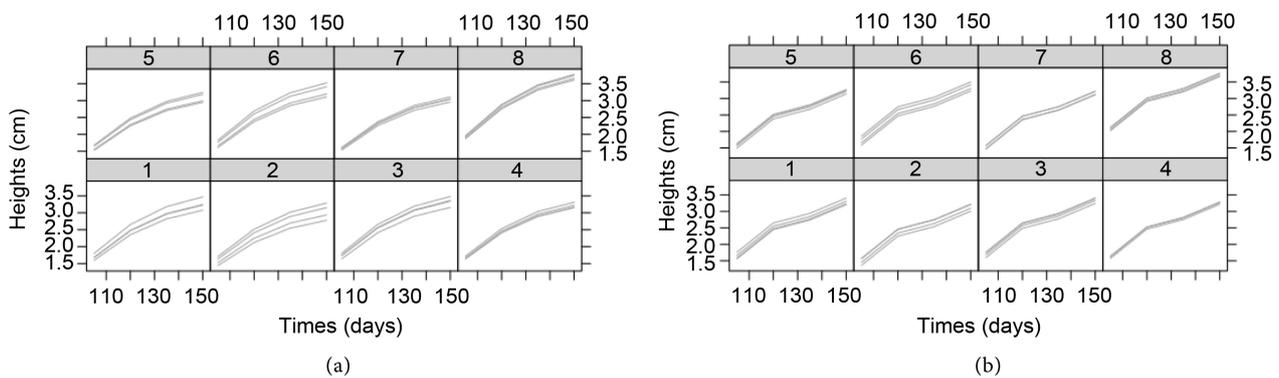(a)   (b)

Figure 8. Fitted growth curves of diameters for each treatment. The sequence of the treatment begins from bottom left (Treatment 1) to top right (Treatment 8). (a) Nonlinear mixed-effect model (2). (b) Nonlinear BCN mixed-effect model (4).
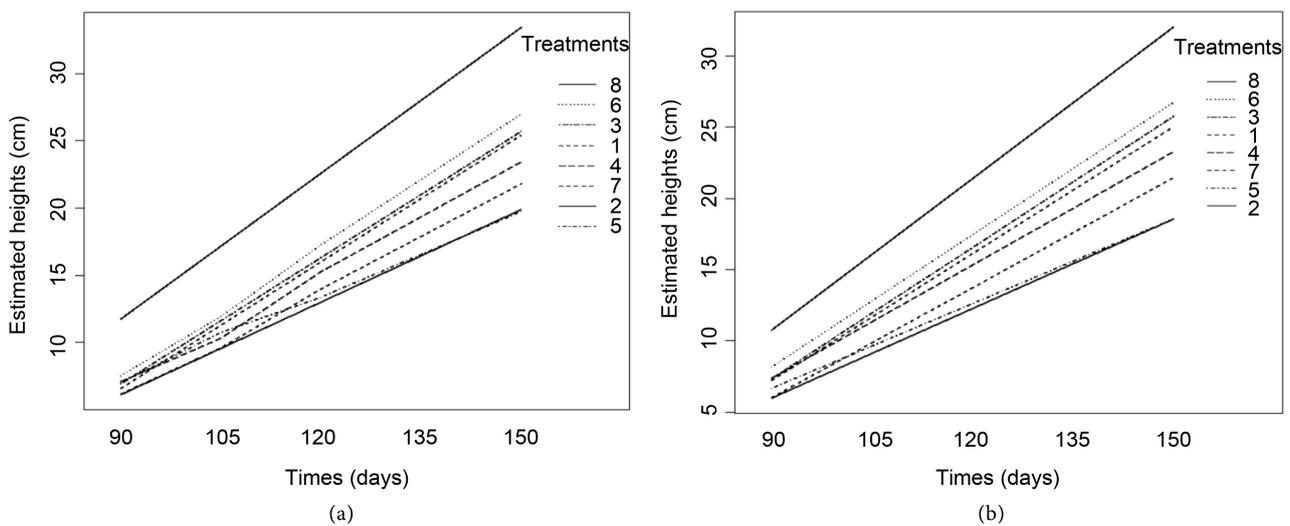

(a)   (b)

Figure 9. Mean profile for the height data. (a) Linear mixed-effect model (1). (b) Linear BCN mixed-effect model (3).
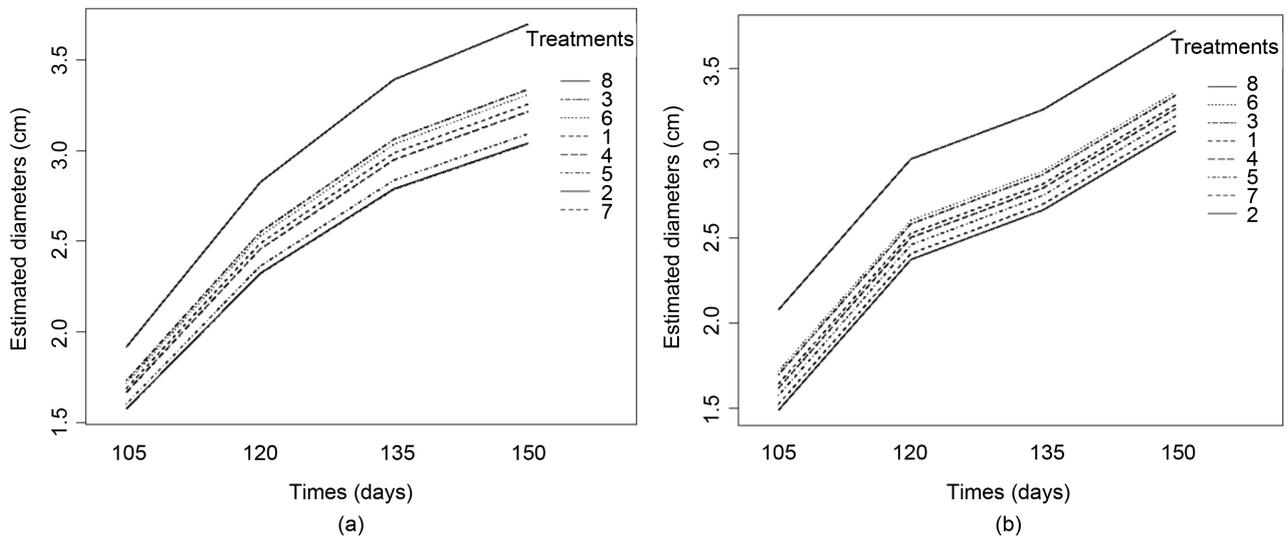
Figure 10. Mean profile for the diameter data. (a) Nonlinear mixed-effect model (2). (b) Nonlinear BCN mixed-effect model (4).

to analyse growth data. Additionally, the classical linear and nonlinear models were used to compare the new proposal. We observed that the fitted models presented similar performances, which means that the BCN model can be used to estimate growth curves. Moreover, its structure considered the slight asymmetry of the data and the AIC criteria primed as the best models. Furthermore, in the presence of outliers, this model can be more proper, because it involves a robust process of estimation. However, more studies can be made involving smooth functions in the nonlinear case.

The growth curves are important for determining the time of delivery of the seedlings to the field. Besides that, the genus *Eucalyptus* species is very important for the economic wood production in Brazil. Finally, curve predictions can be useful when making efficient decisions on the use of this natural renewable resource.

References

Bazzo J. F. (2009). *Use of Organic Sewage Sludge Compost as Substrate for Eucalyptus Seedlings Production*. Master's Thesis, Botucatu (BRA): University of São Paulo State.

Bertalanffy, L. V. (1951). General System Theory: A New Approach to Unity of Science (Symposium). *Human Biology, 23,* 303-361.

Carneiro, J. G. A. (1995). *Production and Quality Control of Forest Seedlings.* https://www.bdpa.cnptia.embrapa.br/consulta/

Carvalho, L. R., Pereira, G. M. S., Silva, H. O. F., Mischan, M. M., & Furtado, E. L. (2014). Nonlinear Models Adjustment of Fixed Effects, with Weight and Mixed-Applications. *Biometric Brazilian Journal, 32,* 296-307.

Cole, T., & Green, P. J. (1992). Smoothing Reference Centile Curves: The LMS Method and Penalized Likelihood. *Statistics in Medicine, 11,* 1305-1319. https://doi.org/10.1002/sim.4780111005

Dias, A. N., Leite, H. G., Nogueira, G. S., & Rodrigues, F. L. (2005). Evaluation of Methods of Site Index Curve Ajustment in Thinned Eucalypt Stands. *Brazilian Journal*

of Forest Science, 29, 741-747.

Ferrari, S. L. P., & Fumes, G. (2017). Box-Cox Symmetric Distributions and Applications to Nutritional Data. *Advances in Statistical Analysis, 101,* 321-344. https://doi.org/10.1007/s10182-017-0291-6

Gomes, J. M. (2001)*. Morphologic Parameters in Evaluting the Quality of Eucalyptus Grandis Seedlings Produced in Different-Sized Tubes and Doses of N-P-K.* PhD Thesis, Viçosa, Brazil: Federal University of Viçosa.

Gomes, J. M., Paiva, H. N., & Couto, L. (1996). Eucalyptus Seedlings Production. *Agriculture Report, 18,* 15-22.

Gomes, J. M., Couto, L., Leite, H. G., Xavier, A., & Garcia, S. L. R. (2002). Morphological Parameters Quality for the Evaluation of Eucalyptus Grandis Seedling. *Brazilian Journal of Forest Science, 26,* 655-664.

Machado, S. A. (1979). Pinus Taeda Survival Estimate in Homogeneous Stands. *Forest Journal, 10,* 73-76.

Machado, S. A., Oliveira, E. B., Carpanezzi, A. A., & Bartoszeck, A. C. P. S. (1997). Site Classification for Mimosa Scabella in the Curitiba Metropolitan Region. *Forest Research Bulletin, 35,* 21-37.

Melesse, S. F., & Zewotir, T. (2017). Variation in Growth Potential between Hybrid Clones of Eucalyptus Trees in Eastern South Africa. *Journal of Forestry Research,* 1-11. https://doi.org/10.1007/s11676-017-0400-0

Oliveira, M. L. R., Leite, H. G., Nogueira, G. S., Garcia, S. L. R., & Souza, A. L. (2008). Classification of the Productive Capacity of Unthinned Stands of Eucalypt Clones. *Brazilian Agricultural Research, 43,* 1559-1567.

Pinheiro, J. C., & Bates, D. (2000). *Mixed Effects Models in S and S-PLUS* (528 p.). New York, NY: Springer. https://doi.org/10.1007/978-1-4419-0318-1

Prodan, M. (1968). *Forest Biometrics* (Gardiner S H, Trans., 447 p.) New York, NY: Pergamon Press.

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* Vienna. https://www.r-project.org/

Richards, F. J. (1959). A Flexible Growth Function for Empirical Use. *Journal of Experimental Botany, 10,* 290-301. https://doi.org/10.1093/jxb/10.2.290

Rigby, B., Stasinopoulos, M., Heller, G., & Voudouris, V. (2014). *The Distribution Toolbox of GAMLSS.* https://www.gamlss.org/

Rigby, R. A., & Stasinopoulos, D. M. (2006). Using the Box-Coxt Distribution in GAMLSS to Model Skewness and Kurtosis. *Statistical Modelling, 6,* 209-229. https://doi.org/10.1191/1471082X06st122oa

Schumacher, F. X. (1939). A New Growth Curve and Its Applications to Timber Yield Studies. *Journal of Forestry, 37,* 819-820.

Silva, J. A. A. (1986). *Dynamics of Stand Structure in Fertilized Slash Pine Plantations.* PhD Thesis, Athens: University of Georgia.

Soares, A. V., Leite, H. G., Cruz, J. P., & Forrester, D. I. (2017). Development of Stand Structural Heterogeneity and Growth Dominance in Thinned Eucalyptus Stands in Brazil. *Forest Ecology and Management, 384,* 339-346.

Stasinopoulos, D. M., Rigby, R. A., & Akantziliotou, C. (2008). *R: A Language and Environment for Statistical Computing Instructions on How to Use the GAMLSS Package in R.* https://www.gamlss.org/

Teo, S. J., Bressan, D. R., & Costa, R. H. (2011). Use of Statistical Models for Site

Classification of *Pinus taeda* Plantation in the Region of Caçador, Santa Catarina State, Brazil. *Forest, 41,* 179-188.

Vanegas, L. H., & Paula, G. A. (2015). A Semiparametric Approach for Joint Modeling of Median and Skewness. *Test, 24,* 110-135. https://doi.org/10.1007/s11749-014-0401-7

Voudouris, V., Gilchrist, R., Rigby, R. A., & Sedgwick, J. (2012). Modelling Skewness and Kurtosis with BCPE Density in GAMLSS. *Statistical Modelling Journal of Applied Statistics, 39,* 1279-1293. https://doi.org/10.1080/02664763.2011.644530