CIÊNCIA e NATURA

UFSM

Ci. e Nat., Santa Maria v.42, e15, p. 01-10, 2019
DOI:10.5902/2179460X41710
ISSN 2179-460X

Meio Ambiente

Submissão: 26/12/2019 Aprovação: 14/02/2020 Publicação: 15/05/2020

Artigo Original

An Application of Generalized Additive Models of Location, Scale, and Shape (GAMLSS) to estimate the Eucalyptus Height

Uma aplicação de modelos aditivos generalizados de locação, escala e forma (GAMLSS) para estimar a altura do Eucalipto

Tiago Almeida Oliveira[I]; Sílvio Fernando Alves Xavier Júnior[II]; Gláucia Amorim Faria[III]; Beatriz Garcia Lopes[IV]; Ednário Mendonça Barbosa[V] e Ana Patrícia Bastos Peixoto[VI]

Abstract

The Generalized Additive Models for Location, Scale, and Shape (GAMLSS) are a recent class of models that further flexibility the distribution of the response variable. The regression analysis has been used to model biological phenomena, and its various modalities have met the need for its use with precision. However, there are situations in which the adjustment of models with more flexible assumptions in the specification of the distribution of the response variable becomes indispensable, thus justifying the useof GAMLSS. The study of plant growth curves has full application in agricultural research; thus, it is crucial to know the habits of growth and development of forest species is crucial for reforestation programs and in the most diverse researches. The study aimedto model the growth of Eucalyptus through the adjusting of Generalized Additive Models for Location, Scale, and Shape, in order to promote improvements on crop productivity. Considering all parameters of the independent variable (time) under GAMLSS class modeling, the distribution model ST3 presented better results.
Keywords: Growth curves; Eucalyptus; GAMLSS; Probabilistic models

Resumo

Os Modelos Aditivos Generalizados para Locação, Escala e Forma (GAMLSS) são uma recente classe de modelos que flexibiliza ainda mais a distribuição da variável resposta. A análise de regressão é muito utilizada para modelar fenômenos biológicos e suas diversas modalidades têm atendido a necessidade do seu uso com precisão. Porém, existem situações nas quais se torna indispensável o ajuste de modelos com pressupostos mais flexíveis na especificação da distribuição da variável resposta. Os Modelos Aditivos Generalizados para Locação, Escala e Forma (GAMLSS) são uma recente classe de modelos que auxiliam o pesquisador a entender o comportamento de fenômenos biológicos, tornando a distribuição da variável resposta mais flexível. O estudo das curvas de crescimento de plantas tem ampla aplicação em pesquisas na área agropecuária, dessa forma, conhecer os hábitos de crescimento e desenvolvimento de espécies florestais é crucial para programas de reflorestamento e nas mais diversas pesquisas. O objetivo do estudo foi modelar o crescimento do eucaliptos, por meio do ajuste de Modelos Aditivos Generalizados para Locação, Escala e Forma, afim de promover melhorias. Considerando todos os parâmetros com efeito da variável independente (tempo) sob a modelagem da classe GAMLSS, o modelo de distribuição ST3 foi o que apresentou melhores resultados.
Palavras-chaves: Curvas de Crescimento; Eucalypto; GAMLSS; Modelos Probabilísticos; Curtose

[I] Universidade Estadual da Paraíba - UEPB- Brasil. tadolive@cct.uepb.edu.br

[II] Universidade Estadual da Paraíba - UEPB- Brasil. profsilviofernando@gmail.com

[III] Universidade Estadual Paulista- Ilha Solteira- UNESP - SP- Brasil. glauciaamorimfaria@gmail.com

[IV] Universidade de São Paulo- USP - SP- Brasil. biagarcialopes@gmail.com

[V] Universidade Estadual da Paraíba - UEPB- Brasil. eddybarbosa92@gmail.com

[IV] Universidade Estadual da Paraíba - UEPB- Brasil. patricia_ufba@yahoo.com.br

1 Introduction

Knowing the habits of growth and development of forest species is crucial in various research, especially in reforestation programs. For eucalyptus cultivation, the study of growth curves will make it possible to define the time ideal for cutting or rotating, which corresponds to the time when the maximum volumetric and economic productivity of a forest stand Gonçalves et al. (2017).

The study of plant growth has a widespread application in agricultural research because it allows the researcher comparing the behavior of the plant or its relevant components are given the various particularities of each experiment Maia et al. (2009). The objective of modeling growth data is to obtain information about the parameters in which the curve is composed and to get more concrete interpretations of the parameters, and in such a way that it is possible to construct a standard model for the studied observations Lopes (2019).

About modeling techniques, regression analysis has been used to model biological phenomena, and its various modalities have met the need for its use with precision. Hereupon, Fumes et al. (2017) address a set of models for height and diameter growth using linear and nonlinear models from many authors, including Schumacher (1939), Von Bertalanffy (1951) Richards (1959), Prodan (2013), known as Brody, Von Bertalanffy, Logistics and Gompertz have been successfully used in many works. These models assume a normal distribution for the set of data, that generally, this assumption is not verified.

When situations arise where there is a need to adjust models with more flexible assumptions in the specification of the response variable distribution. Zhang et al. (2015) discuss that a new class of univariate regression models has to gain space in the most diverse models, as compared to the generalized linear models (GLM) Nelder e Wedderburn (1972) and the classic generalized additive models (GAM) HASTIE (1990) and mixed models, provides a structure of more flexible modeling. The Generalized Additive Model for Location, Scale and Shape (GAMLSS) was developed by Rigby e Stasinopoulos (2005) for non-stationary modeling. Thus, variables of interest may follow a more general distribution than the exponential family (Gaussian and exponential), such as highly distorted or kurtotic distributions, which may be more appropriate for modeling the records of interest. In addition, GAMLSS enables all conditional distribution parameters to be modeled as additive parametric and / or non-parametric (soft) functions of explanatory variables of interest Rigby e Stasinopoulos (2005).

The GAMLSS model class is a robust methodology as it is a powerful tool that allows to model not only the mean, GLM, GAM, but also the dispersion, asymmetry, and kurtosis, generating possibilities for the researcher and, consequently, achieving better adjustments in the modeling process. Thus, this paper aims to apply models to describe the growth of the Eucalyptus by adjusting the GAMLSS Class Models to promote improvements over its culture.

2 Methodology

When fitting models whose distribution of the response variable does not necessarily follow a distribution belonging to the exponential family, we propose to replace them with a family of general distribution. In this sense, Rigby e Stasinopoulos (2005) proposed the class regression models called Generalized Additive Models for Location, Scale and Shape-GAMLSS). GAMLSS is semi-parametric regression models, which assume that for $i = 1, 2, \ldots n$, independent observations y_i have the distribution function $F_y(Y_i|\theta_i)$ where the vector $\theta_T = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{ip})$ represents a vector of the p distribution parameters that represent location, scale, and shape variables. The distribution parameters are related to the selected covariate design matrix using the $gk(\cdot)$ monotonic binding function to $k = 1{,}2, \ldots, p$.

In GAMLSS methodology, the systematic part of the model is expanded to allow modeling not only the mean but all the location, scale, and shape parameters of the response variable distribution. In this study, the linking functions were: identity for the location parameter and logarithm for the other parameters.

A GAMLSS model is described as follows

$$Y \overset{ind}{\sim} D\left(\mu,\sigma,\nu,\tau\right)$$

$$\begin{cases} \eta_1 = g_1\left(\mu\right) = X_1\beta_1 + s_{11}\left(x_{11}\right) + ... + s_{1J_1}\left(x_{1J_1}\right) \\ \eta_2 = g_2\left(\sigma\right) = X_2\beta_2 + s_{21}\left(x_{21}\right) + ... + s_{2J_2}\left(x_{2J_2}\right) \\ \eta_3 = g_3\left(\nu\right) = X_3\beta_3 + s_{31}\left(x_{31}\right) + ... + s_{3J_3}\left(x_{3J_3}\right) \\ \eta_4 = g_4\left(\tau\right) = X_4\beta_4 + s_{41}\left(x_{41}\right) + ... + s_{4J_4}\left(x_{4J_4}\right) \end{cases} \tag{1}$$

where $D\left(\mu, \sigma, \nu, \tau\right)$ is a four-parameter distribution, μ is usually a location parameter, σ is constantly a scale parameter, ν and τ are the shape parameters of the distribution, commonly associated with asymmetry and kurtosis, respectively; η_i represents the linear predictors of each parameter. X_1, X_2, X_3 and X_4 are the arrays that may or may not match, or rather the predictor of each distribution parameter that receives different explanatory variables Rigby e Stasinopoulos (2005).

GAMLSS allows adjustments through linear or nonlinear parametric functions and nonparametric functions. The package `gamlss` allows to smooth additive terms and, together with other packages, allows to adjust decision trees, random effects, neural networks, and multidimensional smoothing simultaneously Stasinopoulos et al. (2017).

The configuration of the distribution $D\left(\mu, \sigma, \nu, \tau\right)$ (1) is general and only implies that the distribution must be in the parametric configuration. The package verb | gamlss | cite stasinopoulos2007generalized implemented exclusively in *software* R, features about 200 discrete, continuous and mixed distributions implemented in the GAMLSS family, integrating some uniquely asymmetric, platicurtic or leptocurtic distributions.

There are three distinct types of distribution families into GAMLSS: continuous, discrete, and mixed distributions. The table with the names of distributions, parametric space, and domain of functions and their binding functions in the implementation of verb | R | can be seen in detail in Stasinopoulos et al. (2017).

2.1 *Skew-t* type 3 distribution

The *Skew-t* type 3 (ST3) distribution was obtained from the *t-student* distribution by the method of Fernandez e Steel (1998), which consists of a transformation of a symmetric distribution into an asymmetric one. Thus, the distribution function (ST3) is given by:

$$f_Z(z;\mu,\sigma,\tau,\nu) = \begin{cases} \kappa\sigma^{-1} \times \left[\left(1+\dfrac{z^2}{\nu}\right)\dfrac{1}{\tau^2}\right]^{-\frac{(\nu+1)}{2}} & , z \geq 0 \\ \kappa\sigma^{-1} \times \left[\left(1+\dfrac{z^2}{\nu}\right)\tau^2\right]^{-\frac{(\nu+1)}{2}} & , z < 0 \end{cases} \tag{2}$$

in which,

$$z = \tfrac{x-\mu}{\sigma}, \kappa = \tfrac{2}{\tau+\frac{1}{\tau}} \tfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{\frac{1}{2}}},$$

where $\mu \in R$ is the location parameter, $\sigma > 0$ scale, $\tau > 0$ form and $\nu > 0$ are the degrees of freedom, respectively. The $ST3$ distribution is one of five asymmetric versions for the t distribution, defined with four parameters, supported by R. The notation, $Z \sim ST3(\mu, \sigma^2, \tau, \nu)$ will be adopted.

Let $y_{11}, \ldots, y_{IJ}$ be a sample size n from the $ST3$ distribution. Then, the logarithm of the likelihood function for the parameter vector $\theta = (\mu, \sigma, \tau, \nu)^T$, under a linear regression model, where:

$$l(\theta) = \sum_{i=1}^{I}\sum_{j=1}^{J} log \left\{ k\sigma^{-1} \times \left[\left(1+\frac{z^2}{\nu}\right)\frac{1}{\tau^2}\right]^{-\frac{(\nu+1)}{2}} \right\} \tag{3}$$

whichin, $z_{ij} = \left(y_{ij} - (\beta_0 + \beta_1 x)\right)/\sigma$.

The maximum likelihood estimates $\widehat{\theta}$ of the parameter vector are obtained by maximizing log-likelihood function. To obtain the components of the score vector $U(\theta)$ from the $ST3$ model, the log-likelihood function (3) was derived from ν, μ, and σ. Thus, the components of the score vector $U(\theta)$ are given by Eloy (2016):

$$U_\nu(\theta) = \sum_{i=1}^{I}\sum_{j=1}^{J} \left\{ \frac{\psi\left(\frac{\nu+1}{2}\right)\sqrt{\pi\nu}\tau^2 + \Gamma\left(\frac{\nu}{2}\right)\left[\psi\left(\frac{\nu+1}{2}\right)\pi\nu - \psi\left(\frac{\nu}{2}\right)\pi\nu - \pi\right]}{2\sqrt{\pi\nu}} \right\} \tag{4}$$

$$+\sum_{i=1}^{I}\sum_{j=1}^{J} \left\{ -\frac{1}{2} ln\left[\frac{\nu\sigma^2 + (y_{ij}-\mu)^2}{\nu\sigma^2\tau^2}\right] + \frac{(\nu+1) + (y_{ij}-\mu)^2}{2\nu\left[\nu\sigma^2 + (y_{ij}-\mu)^2\right]} \right\}, \tag{5}$$

$$\tag{6}$$

$$U_\mu(\theta) = \sum_{i=1}^{I}\sum_{j=1}^{J} \left\{ \frac{\left(\frac{\nu+1}{2}\right)(-2y_{ij} + 2\mu)}{\nu\sigma^2 + (y_{ij}-\mu)^2} \right\}, \tag{7}$$

$$\tag{8}$$

$$U_\tau(\theta) = \sum_{j=1}^{J} \left\{ \frac{-\tau^2 + \Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu} - (\nu+1)\left[\tau^2 + \Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\right]}{\tau\left(\tau^2 + \Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\right)} \right\}, \tag{9}$$

$$\tag{10}$$

$$U_\sigma(\theta) = \sum_{i=1}^{I}\sum_{j=1}^{J} \left\{ \frac{1}{\sigma} - \frac{(\nu+1)(y_{ij}-\mu)^2}{\sigma[\nu\sigma^2 + (y_{ij}-\mu)^2]} \right\}. \tag{11}$$

Equating these equations to zero and solving them the maximum likelihood estimates of the parameters using numerical methods is obtained.

2.2 Model Selection

Selecting models in GAMLSS includes selecting the best distribution for the response variable, the appropriate predictors for the selected distribution parameters, the binding functions, and the hyperparameters.

The Generalized Akaike Criterion (GAIC), used in the same context as the Akaike Information Criterion (AIC) Akaike (1998), takes into account the number of parameters and degrees of freedom used in the model to penalize the most complex models, and avoid overfitting data on large sample sizes Paiva et al. (2008). GAIC is defined by Voudouris et al. (2012) as follows

$$GAIC\,(k) = -2l\left(\hat{\theta}\right) + (k \times gl)\,,$$

where l is the logarithm of the likelihood function and gl are the useful degrees of freedom of the adjusted model, k is constant, and is the penalty for each degree of freedom used, $theta$ refers to the parameters. Refers to $-2l\left(\hat{\theta}\right)$ as the global deviation, because $GAIC\,(k)$ is the statistic obtained by adding the global deviation.

Global deviation (GDEV), an essential measure for selecting models in GAMLSS, is defined as,

$$GDEV = -2l\left(\hat{\theta}\right),$$

wherein, $l\left(\hat{\theta}\right)$ é o logaritmo da função de verossimilhança ajustada.

The model with the lowest value of the $GAIC\,(k)$ criterion, for some chosen k, is then selected and considered the best fit, since $GAIC\,(k)$ penalizes models with many parameters. Criteria (AIC) Akaike (1998) and Bayesian Information Criterion (BIC) Schwarz et al. (1978) are special cases of $GAIC\,(k)$, a case if $k = 2$ implicates on (AIC) and another case if $k = ln(n)$ implies on (BIC) Voudouris et al. (2012).

3 Material and Methods

The experiment was conducted at UNESP/Ilha Solteira Campus, with clones of the hybrid VR3 *Eucalyptus grandis* $\times$ *Eucalyptus urophylla*, named *E. urograndis*, from February to June 2019. The soil used was collected in a layer of 0.0 to $0.20m$ deep, under eucalyptus plantation, located in Três Lagoas-MS (Latitude: 20º 59'S and Longitude 51º 48 'O). The soil was sieved ($4mm$) and placed in seedlings bags with a height of 0.40 m (8 kg of soil). Limestone and NPK (Nitrogen, Phosphorus and Potassium) were incorporated according to the recommendation in use in the region for eucalyptus cultivation ($1.5mg/ha$ of limestone, $0.5mg/haNPK$ - $12-20-16$). Eucalyptus seedlings (*Eucalyptus*), used as indicators, were donated by FIBRIA Ltda (www.fibria.com.br). Their growth was evaluated by measuring the crop over 0, 2, 4, 6, 12, 24, 36, 48, 60 months after planting.

In this study, we considered a model of the regression model class, but under the assumption that the observations follow a distribution of the GAMLSS model class, as presented in the package `gamlss` of the statistical software R R Core Team (2018). The model used to describe growth is given by

$$Y_{ij} = \beta_0 + \beta_1 x + \sigma\epsilon_{ij}, \tag{13}$$

em que Y_{ij} representa o valor observado em cada tempo i, β_0 é o efeito do intercepto, β_1 é a taxa de crescimento i e $\epsilon_{ij} \sim$ N(0,σ^2) representam os efeitos do fator não controlado do ensaio experimental, com i = 1,..., I e j = 1,..., J, sendo que I denota o número de observações no tempo e J o número de repetições e pode-se também atribuir $\mu_i = \beta_0 + \beta_1 x_i$.

where Y_{ij} represents the observed value at the time i, β_0 is the effect of the intercept, β_1 is the growth rate i and $\epsilon_{ij} \sim$ N $(0, \sigma^2)$ represent the effects of the uncontrolled factor of the experimental assay, with $i = 1, \ldots, I$ and $j = 1, \ldots, J$, where I denotes the number of observations in time and J the number of repetitions, and one can also assign $\mu_i = \beta_0 + \beta_1 x_i$.

In order to perform the analysis, the packages `gamlss` and `fitDist` were used, both from R R Core Team (2018). The GAIC measure was applied to evaluate the best model with k= 1. Thus, the residual and worm plot diagnostic graphs were displayed. The Shapiro e Wilk (1965) test was used to evaluate GAMLSS model residues as well as the statistics of the first four moments of the distribution under study.

4 Results and Discussion

Figure 1 exhibits the scatterplot over time for height values (cm). It shows the existing relationship, demonstrating good model adjustments for data between eucalyptus plant height over the months, and it suggests a normal growth curve. This curve shows that growth presents a faster acceleration phase from 10 to 40 months, reaching a stabilization phase for commercial cutting around 60 months, which is the period in which the harvest is generally determined because it is when the forest supplies the product in a pertinent quality and quantity for economic return.

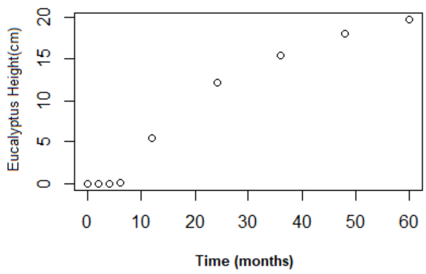


Figure 1: Scatterplot of Time (month) *versus* Eucalyptus Height(cm).

Figure 2a displays a parabolic curve, showing a possible lack of higher terms in the model under the assumption of normality. In this way, even by adding polynomial terms to the model, the adjustment did not improve with the behavior of the residuals with greater deviations from normality. In this case, this parabolic behavior of the residuals may suggest a lack of higher adjustment order (asymmetry and/or kurtosis). GAMLSS models can be used in this case, because in these situations where they have asymmetrical behaviors that traditional modeling cannot correct Stasinopoulos et al. (2007). Thus, data adjustments were made for a variety of probabilistic models of the GAMLSS class in order to choose the one that best characterized the phenomenon. In all, 60 continuous distributions were adjusted, of which the 30 best are presented in table 1.

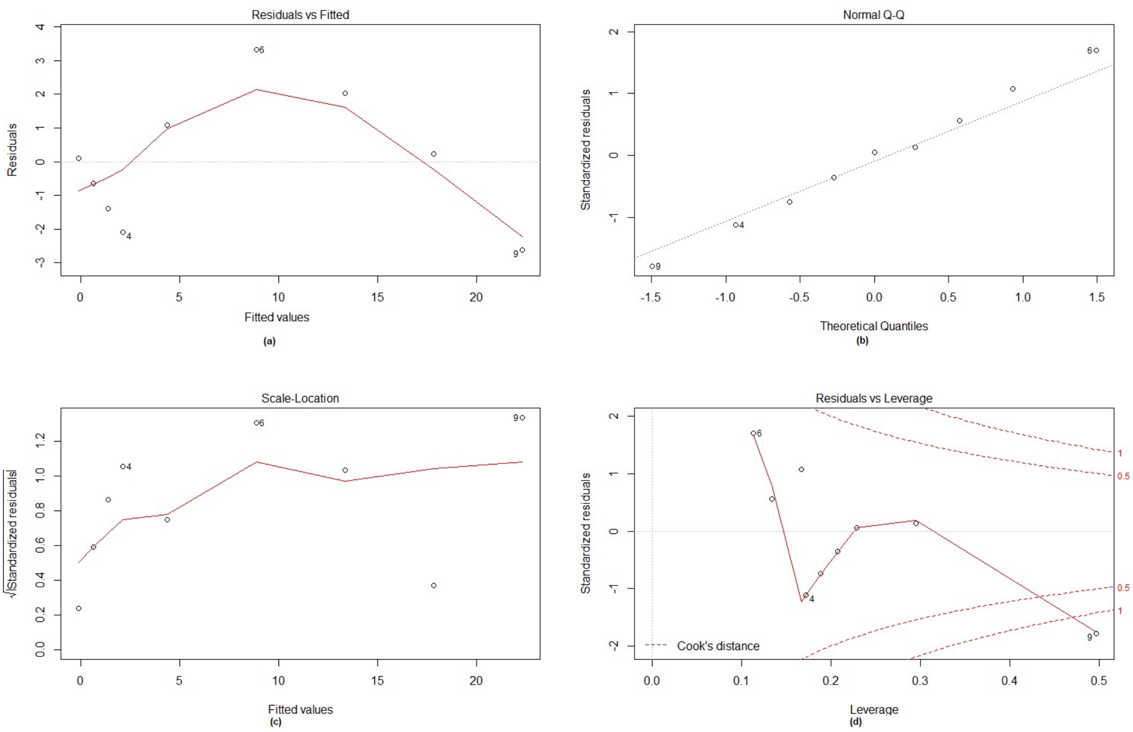


Figure 2: Plots of Residuals for Eucalyptus height.

The classification of the distributions was performed via the lowest GAIC value. The acronyms related to the distributions cited in table 1 can be further detailed as well to their meaning and their statistical properties in Stasinopoulos et al. (2017).

Table 1: Generalized Akaike Information Criterion for Probabilistic Models Adjusted via gamlss Package.

Distrib.	GAIC	Distrib.	GAIC	Distrib.	GAIC	Distrib.	GAIC
SHASH	29.79366	LOGNO	40.82479	PARETO2o	42.52187	SN2	62.56368
GG	35.34865	LOGNO2	40.82479	GP	42.52187	ST3	64.56368
GIG	38.01075	BCCGo	41.12412	BCTo	43.12412	ST2	64.56370
GA	38.02769	BCCG	41.12412	JSU	48.39924	ST1	64.56402
WEI	39.86231	IGAMMA	41.35272	IG	49.21181	SST	64.66178
WEI2	39.86231	GB2	42.27532	EXP	57.15536	RG	66.10766
WEI3	39.86231	PARETO2	42.52187	SEP1	61.48246	NO	66.91545

The *Sinh-Arcosinh*-SHASH distribution presented the lowest value via GAIC. However, when the linear effect of time was added (Figure 4), both the SHASH distribution and the rest of the list do not fit satisfactorily when compared to the normal distribution (Table 2 and Figure 3). Due to the proximity of the results between the distributions, it was also adjusted with the *original Sinh-Arcosinh* - SHASHo distribution, but it did not present satisfactory results either (Figure 5). In this sense, the *Skew t* distribution type 3 (ST3) presented improved values of the quantile residual statistics, as well as a better plot of residuals and *worm plot* (Figure 6).

After the inclusion of the fixed effects in the probabilistic models, one can see that the Normal distributions plus the σ (Normal + σ) parameter modeling, Normal, SHASHo, SHASH, and ST3, with different parametric configurations (table reftable2) did not present a better GAIC value in comparison with the SHASHo distribution. On the other hand, the residual quantiles and residual graphical analysis values (Figure 5) were not satisfactory, taking the choice of ST3 distribution, which obtained lower GAIC value than Normal distribution and better residual quantiles values than the other distributions evaluated.

Table 3 shows the results for the mean of the five distributions considered (Normal, Normal $+\sigma$, SHASHo, SHASH, and ST3), which were close to zero, with the most significant difference for the SHASHo and SHASH model. Regarding variance, SHASHo and ST3 were the worst performing distributions, with a difference of 0.20 units. Regarding the Asymmetry Coefficient, Normal $+\sigma$ presents the highest asymmetry followed by Normal and SHASH, with ST3 and SHASHo performing better. For kurtosis, only SHASHo achieved satisfactory performance, and no notable difference can be related to the Filliben Correlation Coefficient.

One can see that the statistics of ST3 were one of the best, with one of the lowest values of asymmetry and when analyzed together with the graphs of Figure 6, it is observed that the distribution was better suited to the data.

Figure 3: Worm plot and residual graphs of Normal distribution (mean and dispersion).

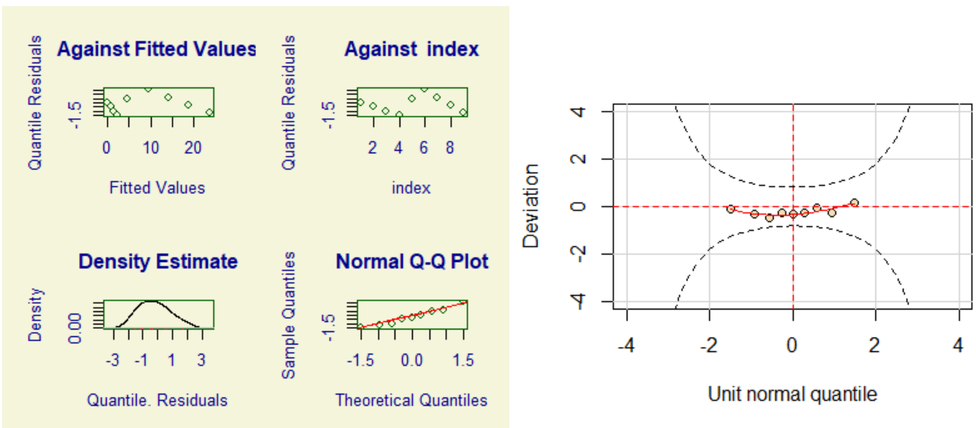


Figure 4: Worm plot and residual graphs of SHASH distribution.

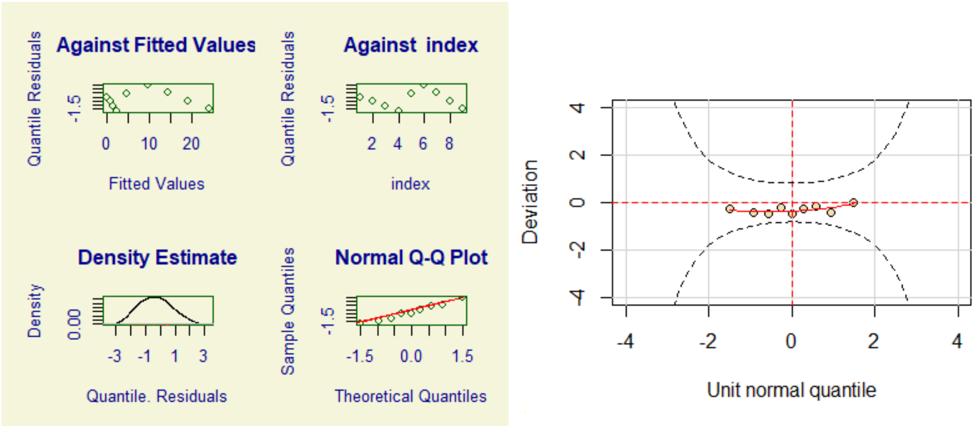
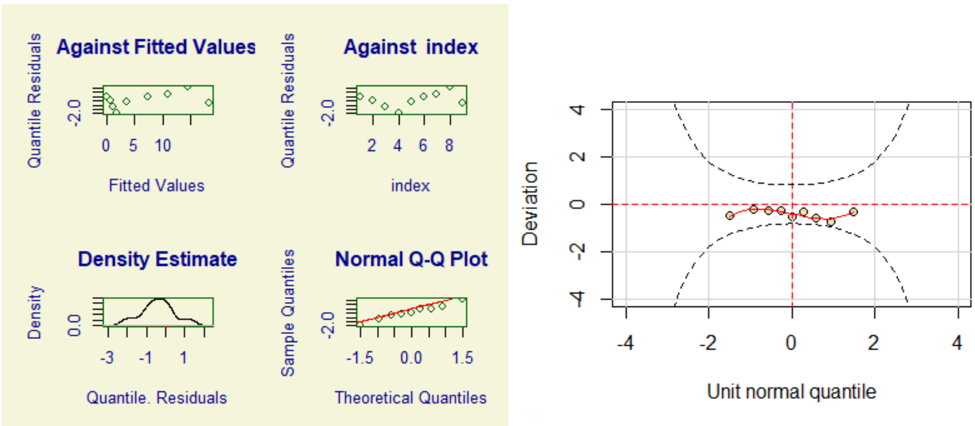


Figure 5: Worm plot and residual graphs of SHASHo distribution.



Regarding the effects of the estimates of the parameters μ, σ, τ, andν, the linear effect of β_1 on growth was 0.54cm for each time unit. This result is different from the regression model under normal residual distribution, where each unit of time would have an average increase of 0.37cm. The other parameters are not interpreted in this work, but we have that, except for τ, the parameters were significant, indicating the need to consider these parameters to model the dispersion, asymmetry and kurtosis arranged in table 4.

Figure 6: Worm plot and residual graphs of ST3 distribution.

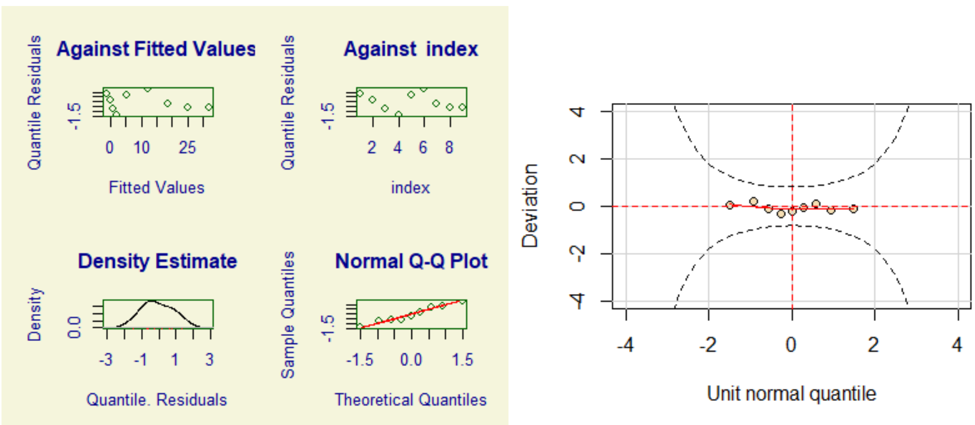


Table 2: Generalized Akaike Information Criterion (GAIC) of probabilistic models adjusted via GAMLSS package in different parametric configurations.

Distribution	GAIC	μ	σ	τ	ν
SHASHo	38.01075	x	x	x	x
SHASH	38.02769	x	x	x	x
Normal+σ	38.73810	x	x		
ST3	39.16977	x	x	x	x
Normal	39.37939	x	$\times$		

$\times$ Intercept; x Covariate effects (Linear (β_1) or Quadratic (β_2)).

Table 3: Residual Quantis Statistics for Proposed Models.

Statistics	Models				
	Normal	Normal + σ	SHASHo	SHASH	ST3
Mean	-0.2337058 $\times 10^{-18}$	-0.214825	-0.4105777	-0.3067954	-0.07357571
Variance	1.125	1.073074	0.8098819	1.058978	0.8001895
Assymetry	0,240712	0.3458527	0.00167738	0.2057842	0.1143107
Kurtosis	1.660571	1.786863	2.265013	1.745959	1.629309
Filliben Correlation Coefficient	0.9918957	0.9856064	0.9819349	0.9897563	0.9849912

Table 4: Statistics of the GAMLSS model adjusted under the *Skew* type 3 distribution.

	Link Function of μ: Coefficients of μ			
	Estimate	Std Error	t-value	Pr(>|t|)
β_0	-1,13	0,86	-1,31	0,26
β_1	0,54	0,05	11,24	0,00
	Link Function of σ: log. Coefficients of σ			
	Estimate	Std Error	t-value	Pr(>|t|)
β_2	-5,28$\times 10^{-6}$	-1,31$\times 10^{-6}$	-401,30	0,00
	Link Function of ν: log. Coefficients of ν			
	Estimate	Std Error	t-value	Pr(>|t|)
β_1	-0,04	0,01	-4,97	0,01
	Link Function of τ: log. Coefficients of τ			
	Estimate	Std Error	t-value	Pr(>|t|)
β_1	0,56	0,98	0,57	0,60

In general, for growth curves and growth models, GAMLSS has been used frequently. Reference curves linked to fetal growth behavior for anthropometric parameter values in the population were used by Paiva et al. (2008).The authors Fumes et al. (2017) studied a class of models for adjusting growth curves for diameter and height of Eucalyptus grandis $\times$ Eucalyptus urophylla: a mixed longitudinal model using a new Box-Cox Normal (BCN) distribution approach on the GAMLSS class, made comparisons with the model assuming normality of the data. According to the AIC criterion, the BCN model presented the best result considering the small asymmetry in the data set.

Thomas et al. (2018) stated that GAMLSS allows different types of random effects to be included in the model's formulation. The authors performed an analysis of an experiment with plant growth data using the GAMLSS methodology with a random effect and compared it to a mixed linear model approach obtaining better results for the GAMLSS model under the Generalized Gamma (GG) distribution.

In the literature there are still few studies involving the GAMLSS model with tree species, neither with Eucalyptus. However, other methodologies have been utilized in the study of growth curves by several authors such as de Mendonça et al. (2014) and Melo et al. (2017) who used the logistic model to estimate wood production and describe plant height, respectively. Using the Gompetz model, Vendruscolo et al. (2015a) described the height and diameter of eucalyptus. Vendruscolo et al. (2015b) with the models of Gompetz, Stoffels, Curtis and Trorey described the height and diameter behavior of the plant. With several models, including Logistic and Gompertz models, Santos et al. (2017) demonstrated growth curves for plant height, diameter, basal area and volume.

5 Conclusion

The use of GAMLSS models was adequate to modelling data of Eucalyptus growth over time, demonstrating the versatility of this methodology in the most diverse areas of statistical application. GAMLSS class models had better results than normal distribution, having at least three competitive models to model growth (SHASHo, SHASH, ST3). The model under ST3 distribution was the best, considering all parameters modeled with effects of the independent variable (time).

The estimates of the parameters μ, adjusted for σ, τ, and ν, considering the linear effect of β_1 on growth was 0.54cm for each time unit. In this sense, the growth of eucalyptus over time, modeled by the theory of generalized location, scale and shape (GAMLSS), as opposed to growth curve models, can be understood as one of the first applications in this area.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. Em: *Selected papers of hirotugu akaike*, Springer, pp. 199–213.

Eloy, R. A. O. (2016). Distribuição assimétrica t-student tipo 3: uma aplicação a delineamentos inteiramente casualizados. .

Fernandez, C., Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, *93*(441), 359–371.

Fumes, G., Demétrio, C. G. B., Villegas, C., Corrente, J. E., Bazzo, J. F. (2017). Growth curves for diameter and height using mixed models: An application in eucalyptus seedling. *Open Journal of Forestry*, *7*(04), 403.

Gonçalves, J. C., de Oliveira, A. D., de Pádua Chaves, S., Gomide, L. R., et al. (2017). Análise económica da rotaçâo florestal de povoamentos de eucalipto utilizando a simulaçâo de monte carlo. *Ciência Florestal*, *27*(4), 1339–1347.

HASTIE, R., T. ; TIBSHIRANI (1990). Generalized additive models. [s.1.]. *Wiley Online Library*.

Lopes, B. G. (2019). Classificação da variação, tamanho ótimo de parcela e curva de crescimento para experimentos com eucalipto. Dissertação de Mestrado, Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira.

Maia, M. C. C., Resende, M. D. V., de Paiva, J. R., Cavalcanti, J. J. V., de Moura Barros, L. (2009). Seleção simultânea para produção, adaptabilidade e estabilidade genotípicas em clones de cajueiro, via modelos mistos. *Pesquisa Agropecuária Tropical*, *39*(1), 43–50.

Melo, E., Calegario, N., Mendon A., Possato, E., Alves, J., Junior, M. (2017). Modelagem não linear da relaç ão hipsometrica e do crescimento das arvores dominantes e codominantes de Eucalyptus sp. *Ciência Florestal*, *27*(4), 1325-1338.

de Mendonça, A. R., Calegario, N., da Silva, G. F., de Souza, A. L., Trugilho, P. F., e Carvalho, S. P. C., Possato, E. L. (2014). Modelagem da produção de sortimentos em povoamentos de eucalipto. *CERNE*, *20*(4), 587-594.

Nelder, J. A., Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384.

Paiva, C. S. M., Freire, D. M. C., Cecatti, J. G. (2008). Modelos aditivos generalizados para posição, escala e forma (gamlss) na modelagem de curvas de referência. *Rev bras ciênc saúde*, *12*(3), 289–310.

Prodan, M. (2013). *Forest biometrics*. Elsevier.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL `https://www.R-project.org/`.

Richards, F. (1959). A flexible growth function for empirical use. *Journal of experimental Botany*, *10*(2), 290–301.

Rigby, R. A., Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(3), 507–554.

Schumacher, F. (1939). A new growth curve and its application to timber yield studies. *Journal of forestry*, *37*(10), 819–20.

Schwarz, G., et al. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Stasinopoulos, D. M., Rigby, R. A., et al. (2007). Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, *23*(7), 1–46.

Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., De Bastiani, F. (2017). *Flexible regression and smoothing: using GAMLSS in R*. Chapman and Hall/CRC.

Thomas, G., Pereira, A. I. d. A., Lobos, C. M. V., et al. (2018). Analysis of a longitudinal multilevel experiment using gamlsss. *arXiv preprint arXiv:181003085*.

Vendruscolo, D., Drescher, R., Souza, H., Moura, J. V. M., Mamore, F., Siqueira, T. (2015a). Estimativa da altura de eucalipto por meio de regressão não linear e redes neurais artificiais. *Revista Brasileira de Biometria*, *33*(4), 556-569.

Vendruscolo, D., Drescher, R., Souza, H., Silva, R. (2015b). Estratificação hipsometrica em plantios de eucaliptos na região sudeste de Mato Grosso. *Agrarian Academy*, *2*(3), 52-61.

Von Bertalanffy, L. (1951). General system theory, a new approach to unity of science. 5. conclusion. *Human biology*, *23*(4), 337.

Voudouris, V., Gilchrist, R., Rigby, R., Sedgwick, J., Stasinopoulos, D. (2012). Modelling skewness and kurtosis with the bcpe density in gamlss. *Journal of Applied Statistics*, *39*(6), 1279–1293.

Zhang, D. d., Yan, D. h., Wang, Y. C., Lu, F., Liu, S. h. (2015). Gamlss-based nonstationary modeling of extreme precipitation in beijing–tianjin–hebei region of china. *Natural Hazards*, *77*(2), 1037–1053.