



Nonlinear growth models: An alternative to ANOVA in tomato trials evaluation



Bruno Giacomini Sari*, Alessandro Dal'Col Lúcio, Cinthya Souza Santana, Tiago Olivoto, Maria Inês Diel, Dionatan Ketzer Krysczun

Federal University of Santa Maria, Departament of Plant Science, Av. Roraima 1000, 97105-900, Santa Maria, RS, Brazil

ARTICLE INFO

Keywords:

Solanum lycopersicum
Nonlinear regression
Productive precocity
Productive behavior
Logistic model
Repeated measures

ABSTRACT

The aim of this work was to use growth models as an alternative analysis of experiments with tomatoes. The data were obtained from field experiments of tomatoes carried out in the 2015/2016 and 2016/2017 growing seasons. Six and nine harvests were carried out in 2015/2016 and 2016/2017, respectively. In each harvest, the number and mass of fruits per plant were assessed. The ANOVA for repeated measures was performed to analysis of the data, and the residuals were assessed. Also for each variable, the Brody, Gompertz, Logistic and von Bertalanffy growth models were fitted as a function of the accumulated values per plant in each harvest and days after the transplant of the seedlings. The assumptions of normality, homogeneity and independence of residuals were verified by the Shapiro-Wilk, Breusch-Pagan and Durbin-Watson tests, respectively. The bias of the parameters was evaluated by the intrinsic and parametric nonlinearity, by the difference between OLS-based and the mean of 10.000 bootstrap-based parameter estimates, and by evaluating the symmetry of the sampling distributions of such parameters. Among the models tested, the one that presented a high coefficient of determination and less non-linearity was selected. The parameters of the selected model were compared by the F test. From the second-, third- and fourth-order derivatives with respect to x (days after transplant), the maximum acceleration point, inflection point, maximum deceleration point, and point of asymptotic deceleration were estimated. The residuals increased with the mean when ANOVA was used, and the data transformation did not fix the violated assumption. The violation of the model's assumptions motivated us to propose nonlinear models as an alternative to statistical analysis with ANOVAs. The logistic model had parameters close to being unbiased, and all assumptions have been meeting. Due to this, the logistic model was selected. The critical points of this model allowed assessing the productive precocity, crop concentration period and productive behavior of the genotypes during the productive period. Finally, we have shown that growth models may be an interesting alternative to standard ANOVA procedures to analyze experiments with horticultural crops in future studies.

1. Introduction

Multi harvest crops are characterized by having more than one harvest throughout their productive cycle. These crops present a series of characteristics that must be considered when analyzing the results of the research. For example, the harvests are carried out in the same plant more than once, being classified as repeated measures; these crops are also characterized by uneven fruit maturation and high plant variability, commonly resulting in data with overdispersion, zero-inflated data at the beginning and at the end of the harvests, and heteroscedasticity between rows and between harvests (Lúcio et al., 2016a;

Lúcio and Sari, 2017).

To use the harvests (repeated measures) as a fixed factor can lead to a violation of the independence of residuals in the analysis of variance (ANOVA) since the harvests cannot be randomized in the experiment. In addition, the presence of overdispersion and/or zero-inflated data also leads to violation of assumptions such as normality and homoscedasticity of the residuals, making it impossible to evaluate the data through conventional ANOVA (Lúcio et al., 2011). One way to circumvent these problems is to group plants and harvests so that there is a reduction in heterogeneity and the assumptions of ANOVA are met. Therefore, it is common for the analysis of data from multi-harvest

Abbreviations: MAP, maximum acceleration point; IP, inflection point; MDP, maximum deceleration point; ADP, asymptotic deceleration point; DAT, days after transplant of the seedlings; ANOVA, analysis of variance; OLS, ordinary least square

* Corresponding author.

E-mail address: brunosari@hotmail.com (B. Giacomini Sari).

<https://doi.org/10.1016/j.eja.2018.12.012>

Received 7 November 2017; Received in revised form 18 December 2018; Accepted 30 December 2018
1161-0301/© 2019 Elsevier B.V. All rights reserved.

crops to be carried out considering total production only.

By analyzing the total production only, several important portions of information through the crop cycle are lost; consequently, the inferences are narrowed. However, aiming at solving this problem, alternative statistical approaches to ANOVA can be utilized. Nonlinear regressions are particularly appropriate since when accumulating the values of each harvest, it is observed that a sigmoidal curve can be fitted for the total production (Lúcio et al., 2016b, 2015). Among nonlinear regressions, growth models are widely used in research in the field of agricultural sciences (Bi et al., 2012; Carneiro et al., 2014; Goldberg and Ravagnolo, 2015; Helidoniots et al., 2011; Katsanevakis and Maravelias, 2008; Leduc and Goetz, 2009; Mehtätalo et al., 2015; Mendonça et al., 2014; Wubs et al., 2012).

Nonlinear regression models also have assumptions to be satisfied. The most important of this is the linear approximation. Although the models are nonlinear, the estimates of the parameters are performed by Ordinary Least Square, OLS, (the same used in linear models). As the equations systems do not have directly solve, the iterative process is necessary to solve it, and the linear approximation is essential for the parameters estimates are close to being unbiased (Seber and Wild, 2003; Bates and Watts, 2007).

Although of the nonlinearity of the model, the homoscedasticity, normality, and independence of the residuals are necessary for the construction of confidence intervals and hypotheses testing. The homoscedasticity and independence of the residuals can be violated since horticultural crops have high heterogeneity. However, these problems can be overcome by estimating the parameters by using weighted least squares method, or by using autoregressive errors, using the generalized least squares estimation method (Seber and Wild, 2003; Ritz and Streibig, 2008).

In addition to solving the statistical problem in data analysis of horticultural crops, growth models can increase the inferential utility of the results, since these models have parameters and critical points with a biological interpretation. The asymptote of the curve can be used to determine the total production, and the growth parameter can be used to make inferences about the speed (rate) of fruit production. Other points of the growth curve, such as the inflection point and the points of acceleration and deceleration (Mischan et al., 2011; Pinho et al., 2014; Seber and Wild, 2003) can help researchers to make inferences regarding productive behavior over time, the precocity of production and the concentration of the production. According to Regazzi (2003), Seber and Wild (2003), Ritz and Streibig (2008) and Archontoulis and Miguez (2015), the use of dichotomous variables (dummy variables) allows for verifying whether or not a set of models has the same parameter value through the F test, which allows for comparing the parameters of the models.

The common way to contour the violation of ANOVA assumptions is the data transformation (Piepho, 2009), the use of the linear mixed models (Lee et al., 2008) and the use of generalized linear models when the data does not have a normal distribution. Despite many benefits mainly in relation to the biological interpretation of the parameters, the use of nonlinear models for analyzing data from multiple-harvest crops is still incipient. Nonlinear models not only fit multiple-harvest crops data; they also increase the inferences about the productive behavior (or the growth) of the crops over the time.

The benefits of using growth models in agronomic sciences are still little explored. For many years the main barriers for a wide use of nonlinear models were the high computational effort necessary to estimate the model's parameters and the lack of user-friendly modules for fitting nonlinear regression. This issue was overtaken by the development of the statistical packages with high data processing capacity such as *PROC NLIN* in SAS and *nls()* function in R. The use of nonlinear modes for evaluating experiments with horticultural crops is particularly promising, as they can increase the inferences of the results and circumvent statistical shortcomings of other approaches that arise due to characteristics of the crop. In this regard, the present study aims to

verify if the growth models fit the tomato production data (multiple crop harvests) and if they can be used as an alternative statistical tool to ANOVA.

2. Material and methods

2.1. Experimental data

The data were obtained from two tomato field trials carried out in Department of Plant Science of the Federal University of Santa Maria (UFSM), RS, Brazil ($29^{\circ} 43' S$, $53^{\circ} 43' O$, 95 m above sea level) in the 2015/2016 and 2016/2017 growing season (GS). The Cordillera, Ellen, and Santa Clara hybrids were used in the 2015/2016 experiment and the Cordillera and Gaúcho hybrids were used in the 2016/2017 experiment. The transplanting of the seedlings was carried out on October 17, 2015, in the first experiment, and on September 10, 2016 in the second experiment. The spacings between rows and between the seedlings were 1.00 and 0.30 m in the 2015/2016 experiment, and 1.00 and 0.50 m in the 2016/2017 experiment, respectively.

The training was carried out with plastic wires and the plants were conducted using wood single rods in both experiments. Fertilization was carried out based on the soil chemical analysis and according to the recommendation of the Official Network of Soil and Tissue Analysis Laboratories of the States of Rio Grande do Sul and Santa Catarina (ROLAS, 2004). Pest and disease management followed the technical recommendation of the crop.

Harvestings were carried out at 82, 86, 90, 94, 98 and 102 days after transplant (DAT) in 2015/2016 and at 72, 78, 84, 90, 96, 102, 108, 114 and 120 DAT at 2016/2017. The fruits were harvested when they changed from green to reddish and were later packed into bags and taken to the laboratory where they were counted and weighed using a digital scale with accuracy of 0.01 g.

2.2. ANOVA for repeated measures

The trials were analyzed using the randomized block, with eight replicates, in a factorial scheme (genotypes x harvests) with split plots design. The genotypes were randomized on the whole plots being each plot composed by four tomato plants. The harvests were assigned to the subplot. The statistical model used was

$$Y_{ijk} = \mu + \alpha_i + \gamma_k + \eta_{ik} + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

Where μ is the overall mean, α_i is the fixed effect of the genotypes, γ_k is the fixed effect of the blocks, η_{ik} is the whole plot error, β_j is the fixed effect of the harvests, $(\alpha\beta)_{ij}$ is the fixed effect of the interaction between genotypes and harvests and ε_{ijk} is the subplot errors. After adjusting the models, the assumptions were tested through the residuals vs fitted values plots. The data were transformed if the assumptions were violated.

2.3. Using linear mixed models to determine the covariance matrix structure

The traditional ANOVA requires that the residuals are homoscedastic among the factor levels, but it does not always occur. Lee et al. (2008) suggests using a mixed-effect model framework to verify the variance structure of the errors. As this is not our main focus here, we have demonstrated this approach using only the 2015/2016 GS data. The mixed model used was

$$Y_{ijk} = \mu + \alpha_i + \gamma_k + b_j + (\alpha b)_{ij} + \varepsilon_{ijk}$$

Where μ is the overall mean, α_i is the fixed effect of the genotypes, γ_k is the fixed effect of the blocks, b_j is the random effect of the harvests, $(\alpha b)_{ij}$ is the random effect of the interaction between genotypes and harvests and ε_{ijk} is the error. In this model the variances changing across the harvests. We did this by using a variance function implemented in

the *nlme* package (Pinheiro et al., 2018) of the R software.

2.4. Nonlinear models

The mean values of weight and number of fruits per plant, in each harvest, were accumulated within each genotype and the following growth models were fitted: (i) Brody: $Y_i = \beta_1 - (\beta_2 e^{(-\beta_3 t_i)}) + \varepsilon_i$; (ii) Gompertz: $Y_i = \beta_1 e^{(-\beta_2 e^{(-\beta_3 t_i)})} + \varepsilon_i$; (iii) Logistic: $Y_i = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 t_i)}} + \varepsilon_i$; and (iv) Von Bertalanffy: $Y_i = \beta_1(1 - \beta_2 e^{(\beta_3 t_i)})^3 + \varepsilon_i$, where: Y is the dependent variable (number and weight of fruits) accumulated through the DAT; t_i is the time from DAT to harvest; β_1 is the parameter that represents the asymptotic value; β_2 is the parameter that represents the distance between the initial value and the asymptote, β_3 is the parameter related to the growth rate.

The parameters were estimated using the Ordinary Least Squares method (OLS), using the Gauss-Newton algorithm. Later, the Shapiro-Wilk, Breusch-Pagan and Durbin-Watson tests were applied to verify the normality, homogeneity, and independence of the residuals, respectively (Ritz and Streibig, 2008). The goodness-of-fit of each model for each genotype was determined by the coefficient of determination (R^2) and by verifying the nonlinearity of the model, using the Bates and Watts curvature method (Bates and Watts, 2007).

A bootstrap-based parameter estimation was also performed. In this procedure, 10,000 parameter estimates were obtained for each model. Distribution plots of the 10,000 parameter estimates were constructed for each model. The 95% confidence intervals (CI_{95%}) were computed as the difference between 97.5th and 2.5th percentile of the 10,000 parameter estimates. The bias of such parameters was investigated by using: i) the degree of symmetry of the sampling distribution of parameter estimates; and ii) the difference between the average mean of the bootstrap parameter estimates and the OLS estimates. Asymmetric distributions and high differences between mean bootstrap estimates and OLS estimates were related to biased parameters.

The selected model was that the one with lower nonlinearity measures, parameters estimates close to being unbiased and high goodness-of-fit (R^2 close to one).

2.4.1. Comparing the parameters

In order to verify the equality of the parameters of the model with best goodness-of-fit in predicting weight and number of fruits of the evaluated genotypes, we performed model identity tests, constructed with the use of dummy indicator variables (D_j):

$$D_j = \begin{cases} 1, & \text{if the observation belongs to } j \text{ genotypes;} \\ 0, & \text{if else.} \end{cases}$$

Where $j = 1, \dots, k$ is the evaluated genotypes. Our aim with this test was to compare a complete model (ω), which presents different parameters for each genotype, against a reduced model (Ω), which presents common parameters among the genotypes. Taking the logistic model as an example, the complete model would be given by:

$$Y_{ij} = \sum_{j=1}^k \left[\frac{\beta_{1j}}{1 + e^{(\beta_{2j} - \beta_{3j} t_i)}} \right] + \varepsilon_{ij}$$

Where t_i is the DAT of the i -th harvest for $j = 1, 2, \dots, k$, where k indicates that the observation belongs to genotype k . Note that in this case, a parameter is fitted for each genotype (j). A reduced model would be given by:

$$Y_{ij} = \sum_{j=1}^k \left[\frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 t_i)}} \right] + \varepsilon_{ij}$$

In this case, the parameter β_1 is common for all the k -genotypes. Thus, the formulated hypothesis in this example would be then $H_0: \beta_{11} = \beta_{12} = \dots = \beta_{1k}$, vs H_1 : not all β_{1k} are equal. According to this example, for both variables, the following null hypothesis was tested at

0.05 probability error.

$$\begin{array}{lll} H_0^{(1)}: \beta_{11} = \beta_{12} & H_0^{(1)}: \beta_{21} = \beta_{22} & H_0^{(1)}: \beta_{31} = \beta_{32} \\ H_0^{(2)}: \beta_{11} = \beta_{13} & H_0^{(2)}: \beta_{21} = \beta_{23} & H_0^{(2)}: \beta_{31} = \beta_{33} \\ \vdots & \vdots & \vdots \\ H_0^{(n)}: \beta_{1(k-1)} = \beta_{1k} & H_0^{(n)}: \beta_{2(k-1)} = \beta_{2k} & H_0^{(n)}: \beta_{3(k-1)} = \beta_{3k} \end{array}$$

The comparison between complete and reduced models was performed using the F test (Ritz and Streibig, 2008). The F value was calculated as follow:

$$F_{\text{calc}} = \frac{[\text{SS}_{\text{Error}}(\Omega) - \text{SS}_{\text{Error}}(\omega)] / [\text{DF}_{\text{Error}}(\Omega) - \text{DF}_{\text{Error}}(\omega)]}{\text{MS}_{\text{Error}}(\omega)}$$

It was considered that F_{calc} follows an F distribution with $F_{[v, DF_{\text{Error}}(\omega)]}$ under H_0 , where v is the difference between DF_{Error} of the complete and reduced models.

2.4.2. Obtaining the critical points of the selected model

Finally, the coordinates of the critical points were obtained by using the partial derivatives of the selected model in relation to the independent variable (DAT). The inflection point (IP) was obtained by equating the second-order derivative to zero; the maximum acceleration (MAP) and deceleration (MDP) points were obtained by equalizing the third-order derivative of the model to zero; and the asymptotic deceleration point (ADP) was obtained by equating the fourth-order derivative to zero (Mischan et al., 2011). The curves of the model and the derivatives of the first and second order were represented graphically with the purpose of assisting with an intuitive interpretation of the results.

2.5. Software

All analyses were performed using the software R (R Core Team, 2018). The ANOVA for repeated measures were adjusting using the *split2.rbd()* function of the *ExpDes* package (Ferreira et al., 2018). The linear mixed model was fitted with the *lme()* function of the *nlme* package (Pinheiro et al., 2018). The nonlinear models were fitted using the *nls()* function (R Core Team, 2018), and the bootstrap parameters estimates were obtained using the *nlsBoot()* function of the *nlstools* package (Baty et al., 2015). All graphics were constructed using the *ggplot()* function of the *ggplot2* package (Wickham, 2016).

3. Results

3.1. Analysis of variance

The ANOVA assumptions are that the residuals are independent, follow a normal distribution, and are homoscedastic among factor levels. In our case, the absence of the randomization of the harvests in subplots does not guarantee the independence of the residuals. Another common problem in trials with vegetable crops is the variability among harvests due to the uneven maturation of fruits (Lúcio et al., 2016c; Lúcio and Sari, 2017), which may result in heteroscedastic residuals.

It was observed that the residuals increased with mean in both GS (Fig. 1). In this case, the Box-Cox family data transformation can be used to solve the problem (Piepho, 2009). Even when we used the log ($Y + 0.5$) transformation, the behavior of the residuals remained the same (Fig. 2). Therefore, the ANOVA is inadequate in our case, and this motivated us to proposed alternatives for the data analysis.

3.2. Error structure analysis

To characterize the errors structure we used a linear mixed model approach (Lee et al., 2008). This approach was performed with GS 2015/2016 trial only, and aims to demonstrate statistically that the residuals changes among the harvests. Heteroscedastic residuals are

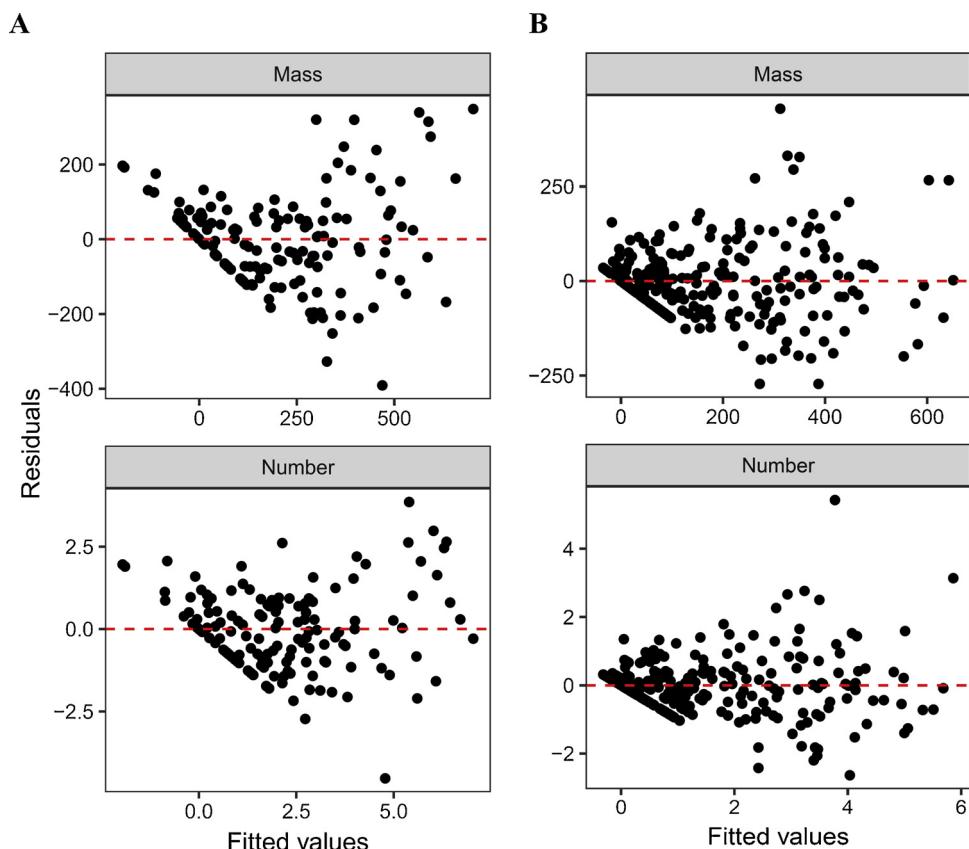


Fig. 1. Residuals vs fitted values plot of the ANOVA in trials conducted in 2015/2016 GS (A) and 2016/2017 GS (B) before data transformation.

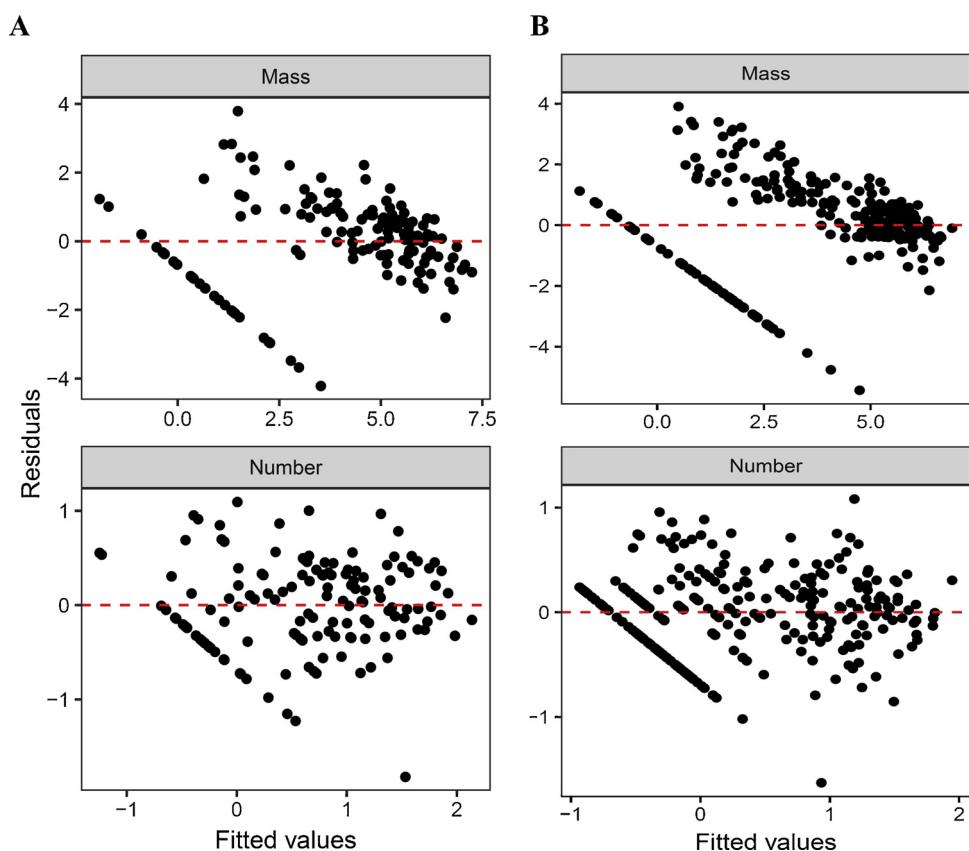


Fig. 2. Residuals vs fitted values plot of the ANOVA in trials conducted in 2015/2016 GS (A) and 2016/2017 GS (B) after data transformation.

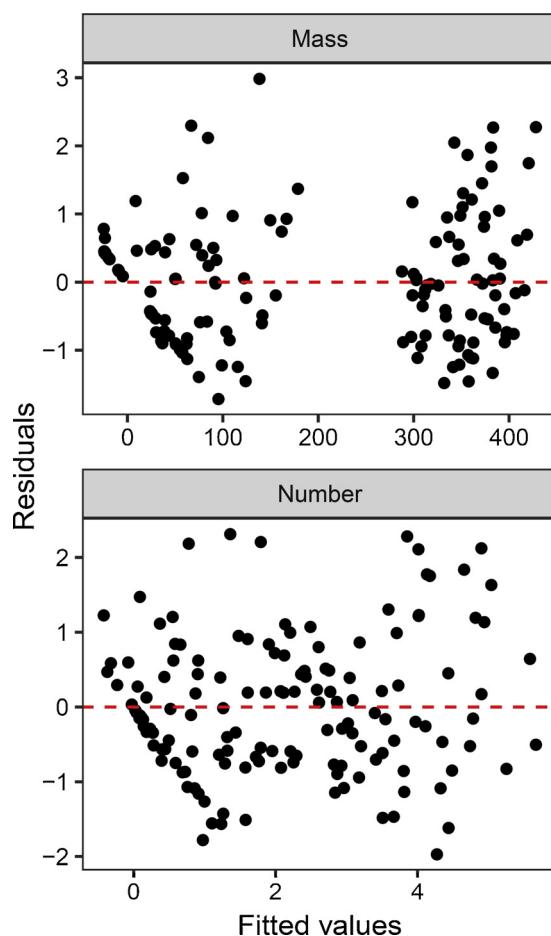


Fig. 3. Residuals vs fitted values plot of the linear mixed model with variance structure that the attributed different variances to the harvests.

very common in vegetable crops trials due to the uneven maturation of fruits (Lúcio et al., 2016c; Lúcio and Sari, 2017), and it must be taken into account in the data analysis. When different variances are attributed to the harvests, we observed that the homoscedasticity of the residuals was meet (Fig. 3).

The intention of this analysis is to demonstrate statistically that the

Table 2

P-values for Shapiro-Wilk (SW), Breusch-Pagan (BP) and Durbin-Watson (DW) tests, determination coefficient (R^2), and nonparametric linearity (c^θ) of the Brody, Gompertz, Logistic and von Bertalanffy models in three tomato genotypes for the number and weight of fruits per plant in 2016/2017 growing season.

Genótipos	Number of fruits					Weight of fruits				
	SW	BP	DW	R^2	c^θ a	SW	BP	DW	R^2	c^θ
Gompertz										
Cordillera	0.34	0.10	0.13	0.99	18.21	0.89	0.12	0.15	0.99	17.57
Gáucho	0.45	0.67	0.52	0.99	19.80	0.74	0.51	0.98	0.99	29.45
Logístico										
Cordillera	0.45	0.46	0.23	0.99	0.81	0.79	0.37	0.28	0.99	0.60
Gáucho	0.08	0.64	0.06	0.99	0.56	0.67	0.87	0.25	0.99	0.50
von Bertalanffy										
Cordillera	0.49	0.14	0.59	0.99	23.24	0.75	0.16	0.61	0.99	15.18
Gáucho	0.94	0.20	0.57	0.99	15.31	0.31	0.21	0.30	0.99	11.94

values obtained by the expression $c^\theta \times \sqrt{F_{(5\%,3,6)}}$.

variances are different among harvests and that when we attribute it through the variance functions, the assumptions of the model are met. The use of the linear mixed models in repeated measures data analysis with variances functions for modeling residuals is widely used in literature (Lee et al., 2008). However, the criticism of the Piepho (2009) regarding the complexity of an analysis with linear mixed models considering several structures of variances motivated us to propose a new analysis tool for multiple crops.

3.3. Models fit and assumptions

All models, with the exception of the Brody model in the 2016/2017 GS, had a good fit to the production data of the evaluated genotypes (R^2 greater than 90%). Thus, the results for the Brody model in the 2016/2017 GS are not presented. It was also observed that the assumptions of normality, homoscedasticity, and independence of residuals were not violated. The logistic model, besides presenting satisfactory goodness-of-fit measures had lower nonlinearity values, i.e., parametric nonlinearity less than 1 (Tables 1 and 2).

3.4. Model selection

The difference between OLS estimate and the mean of the sampling distribution of parameter estimates indicate how much the models have

Table 1

P-values for Shapiro-Wilk (SW), Breusch-Pagan (BP) and Durbin-Watson (DW) tests, determination coefficient (R^2), and nonparametric linearity (c^θ) of the Brody, Gompertz, Logistic and von Bertalanffy models in three tomato genotypes for the number and weight of fruits per plant in 2015/2016 growing season.

Genotypes	Number of fruits					Weight of fruits				
	SW	BP	DW	R^2	c^θ a	SW	BP	DW	R^2	c^θ
Brody										
Cordillera	0.95	0.09	0.71	0.96	444.47	0.99	0.08	0.81	0.97	462.50
Ellen	0.83	0.11	0.90	0.94	485.76	0.95	0.08	0.89	0.96	467.64
Santa Clara	0.06	0.50	0.81	0.98	213.26	0.48	0.33	0.75	0.98	200.06
Gompertz										
Cordillera	0.91	0.07	0.99	0.98	292.06	0.87	0.09	0.91	0.98	301.89
Ellen	0.98	0.09	0.71	0.97	346.58	0.99	0.08	0.91	0.98	316.14
Santa Clara	0.97	0.11	0.19	0.99	95.29	0.89	0.08	0.09	0.99	103.98
Logistic										
Cordillera	0.98	0.07	0.63	0.99	0.58	0.98	0.11	0.84	0.99	0.48
Ellen	0.98	0.08	0.52	0.98	0.97	0.95	0.09	0.61	0.99	0.63
Santa Clara	0.91	0.12	0.88	0.99	0.73	0.87	0.17	0.78	0.99	0.76
von Bertalanffy										
Cordillera	0.83	0.08	0.87	0.98	323.96	0.88	0.09	0.87	0.98	340.26
Ellen	0.94	0.09	0.91	0.97	371.41	0.94	0.08	0.99	0.97	348.42
Santa Clara	0.65	0.22	0.46	0.99	119.67	0.93	0.14	0.29	0.99	121.58

values obtained by the expression $c^\theta \times \sqrt{F_{(5\%,3,3)}}$.

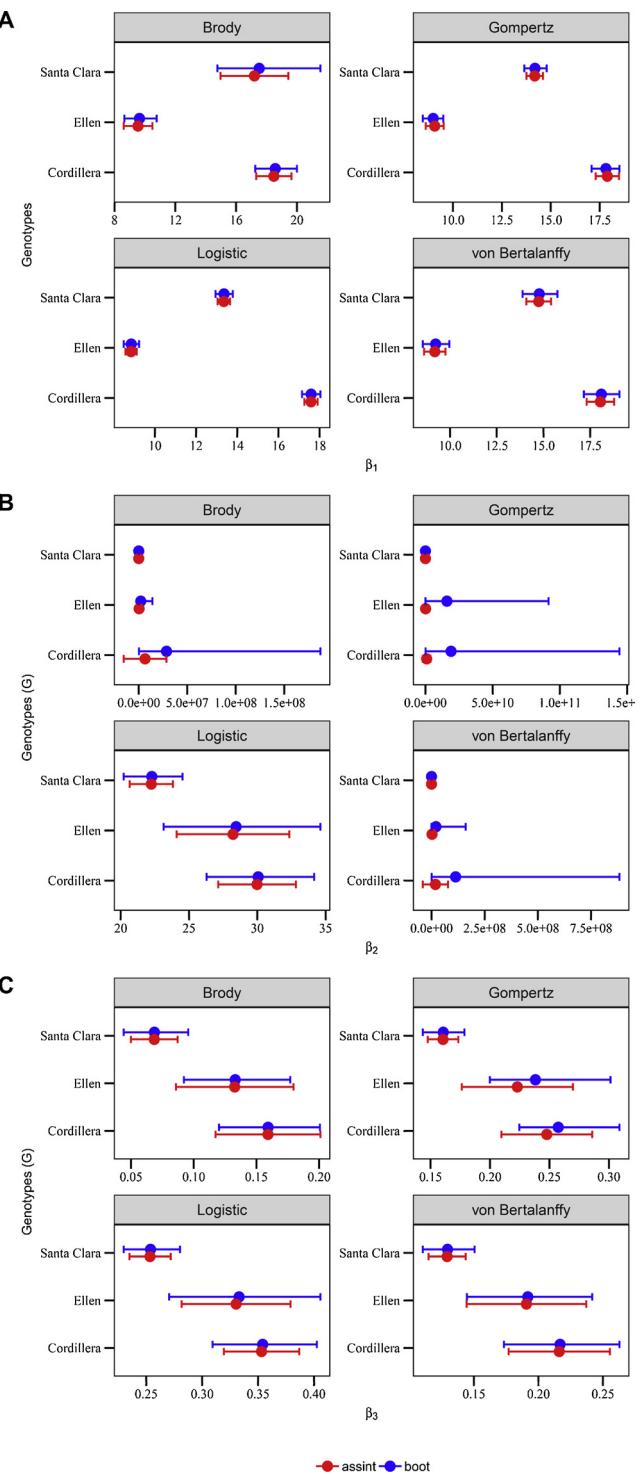


Fig. 4. β_1 (A), β_2 (B) and β_3 (C) estimates and confidence intervals obtained by ordinary least squares (red lines), and Bootstrap-based estimates (blue lines) of the Brody, Gompertz, logistic and von Bertalanffy models fitted to the average number of fruits per plant on 2015/2016 GS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

a good linear approximation and, consequently, the bias such estimates. In GS 2015/2016 (Figs. 4 and 5) the parameter estimates had a large bias than in GS 2016/2017 (Figs. 6 and 7). This is related to the more pronounced sigmoid behavior of the production in GS 2016/2017.

The greater difference between OLS estimates and mean of the sampling distribution of parameter estimates obtained by bootstrap-

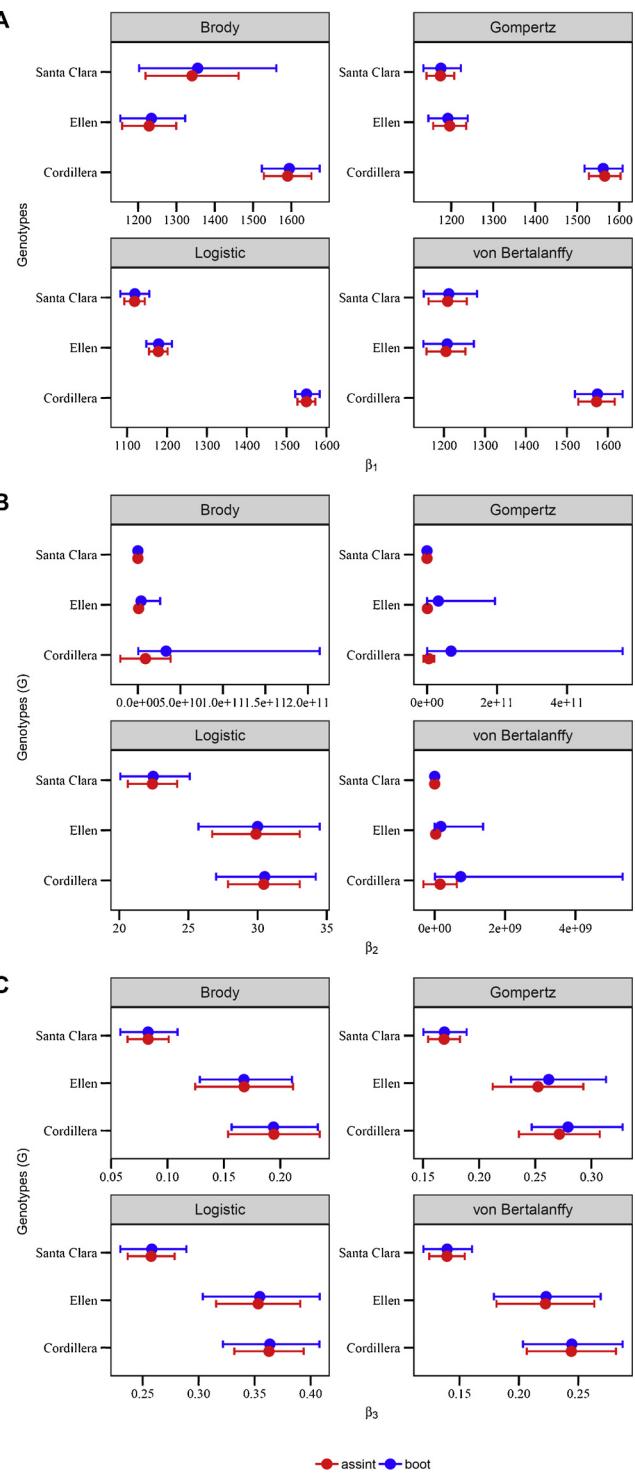


Fig. 5. β_1 (A), β_2 (B) and β_3 (C) estimates and confidence intervals obtained by ordinary least squares (red lines), and Bootstrap-based estimates (blue lines) of the Brody, Gompertz, logistic and von Bertalanffy models fitted to the average mass of fruits per plant on 2015/2016 GS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

approach (Figs. 4–7) and the asymmetrical distribution (Figs. 8–11) indicate that the parameter β_2 is related with the poor linear approximation in Brody, Gompertz, and von Bertalanffy models.

The lower difference between OLS parameters estimates and the mean of the 10,000 parameters estimate obtained by the bootstrap approach, as well as the symmetrical distribution of these estimates are

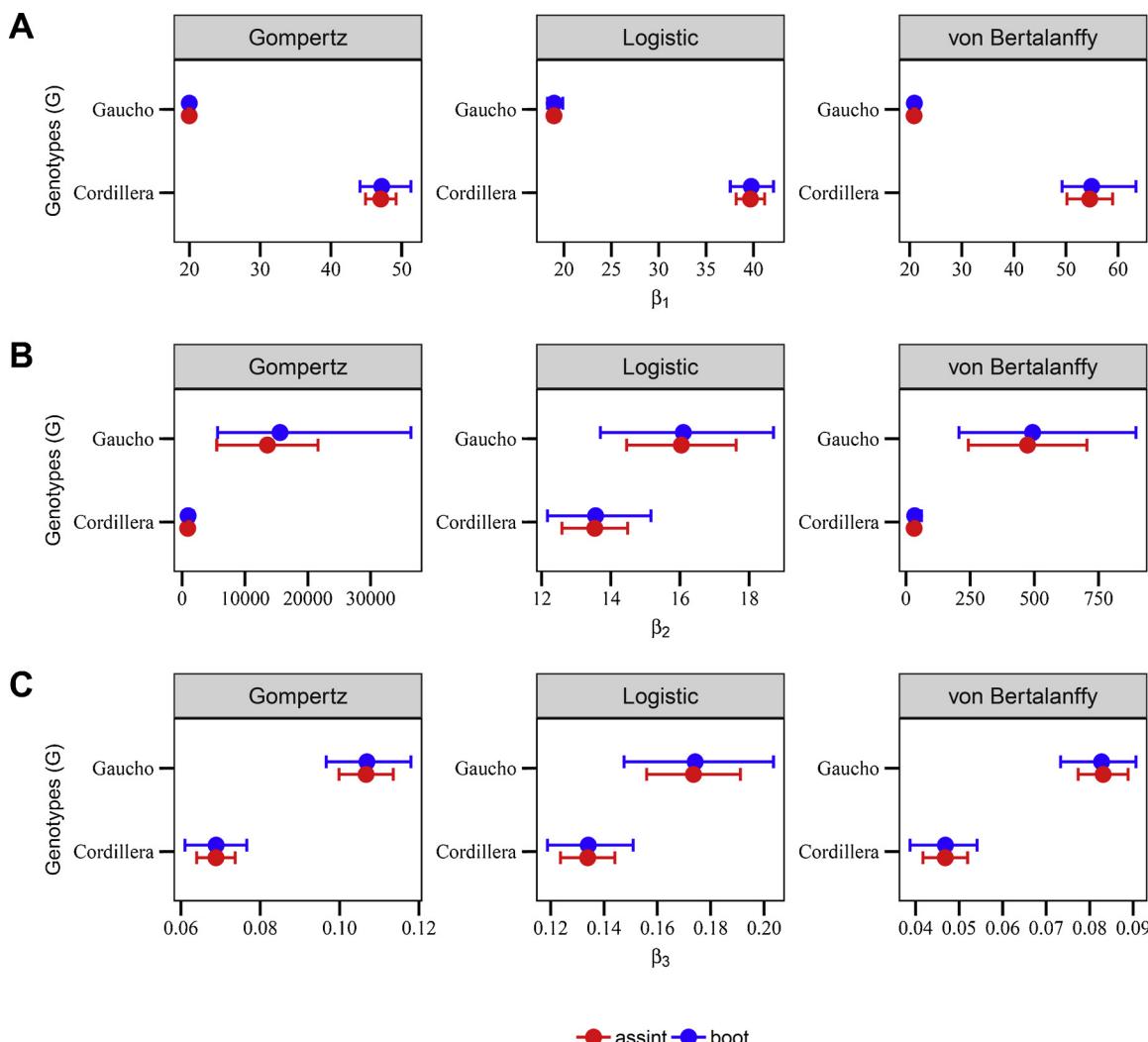


Fig. 6. β_1 (A), β_2 (B) and β_3 (C) estimates and confidence intervals obtained by ordinary least squares (red lines), and Bootstrap-based estimates (blue lines) of the Brody, Gompertz, logistic and von Bertalanffy models fitted to the average number of fruits per plant on 2016/2017 GS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

indicatives that the logistic model has parameter estimates close to being unbiased. It is a direct consequence of the good linear approximation of this model in relation to the other models. The parameter estimates close to being unbiased of the logistic model in more or less sigmoidal patterns, indicate that this model may be used to describe data with forms of productive behavior.

The nonlinear model's selection must be performed based on a linear approximation of the models, and the nonlinearity measures proposed by Bates and Watts are a great alternative to do that. A good linear approximation is related to parameters estimate close to being unbiased, and these parameters are essential to describe the biological phenomena (in our case, the productive behavior) coherently. A poor linear approximation leads to the parameter estimates with a higher degree of bias and, consequently, mistaken conclusions in describing the biological phenomena may occur.

3.5. Describing the productive behavior using the logistic model

The total production of the genotypes can be inferred from the asymptotic value. In the 2015/2016 GS, the Cordillera genotype produced approximately 17 fruits per plant, totaling 1549.72 g per plant. The Ellen genotype produced 8 fruits per plant, $\cong 38\%$ less than the Santa Clara genotype. The average weight per plant was the same, around 1100 g. In the 2016/2017 GS, the Cordillera genotype produced

an average of 39 fruits per plant, with a total weight of 4320.10 g, differing statistically from the Gaúcho genotype, which produced 19 fruits per plant with a total weight of 2921.10 g (Figs. 12 and 13).

The Cordillera genotype had an estimated value of β_2 (scale parameter) equal to the Ellen genotype and superior to the Santa Clara genotype, while the latter two did not differ (Fig. 12). In the 2016/2017 GS, there was no statistical difference between the Cordillera and Gaúcho genotypes for both variables (Fig. 13). Regarding the fruit production rate (β_3), significantly higher values in the Cordillera and Ellen were observed in relation to the Santa Clara genotype in 2015/2016 GS (Fig. 12). In the 2016/2017 GS, the parameters differed statistically only for the weight of fruits, and the rate of production was higher in the Gaúcho genotype (Fig. 13).

The maximum acceleration point (MAP) indicates the moment at which the increase (acceleration) in the rate of fruit production (velocity) is maximum. The beginning of the harvests occurred almost concomitantly with MAP in the 2015/2016 harvest. Therefore, the start was late with plants already producing large amounts of fruits in all genotypes, differently from what occurred in the 2016/2017 harvest, where the harvests started earlier and before reaching MAP (Figs. 14–17).

The inflection point (IP) occurs when the rate of fruit production is maximal, and from it the curve changes from an exponential to a logarithmic trend. In the 2015/2016 GS, the IP was reached

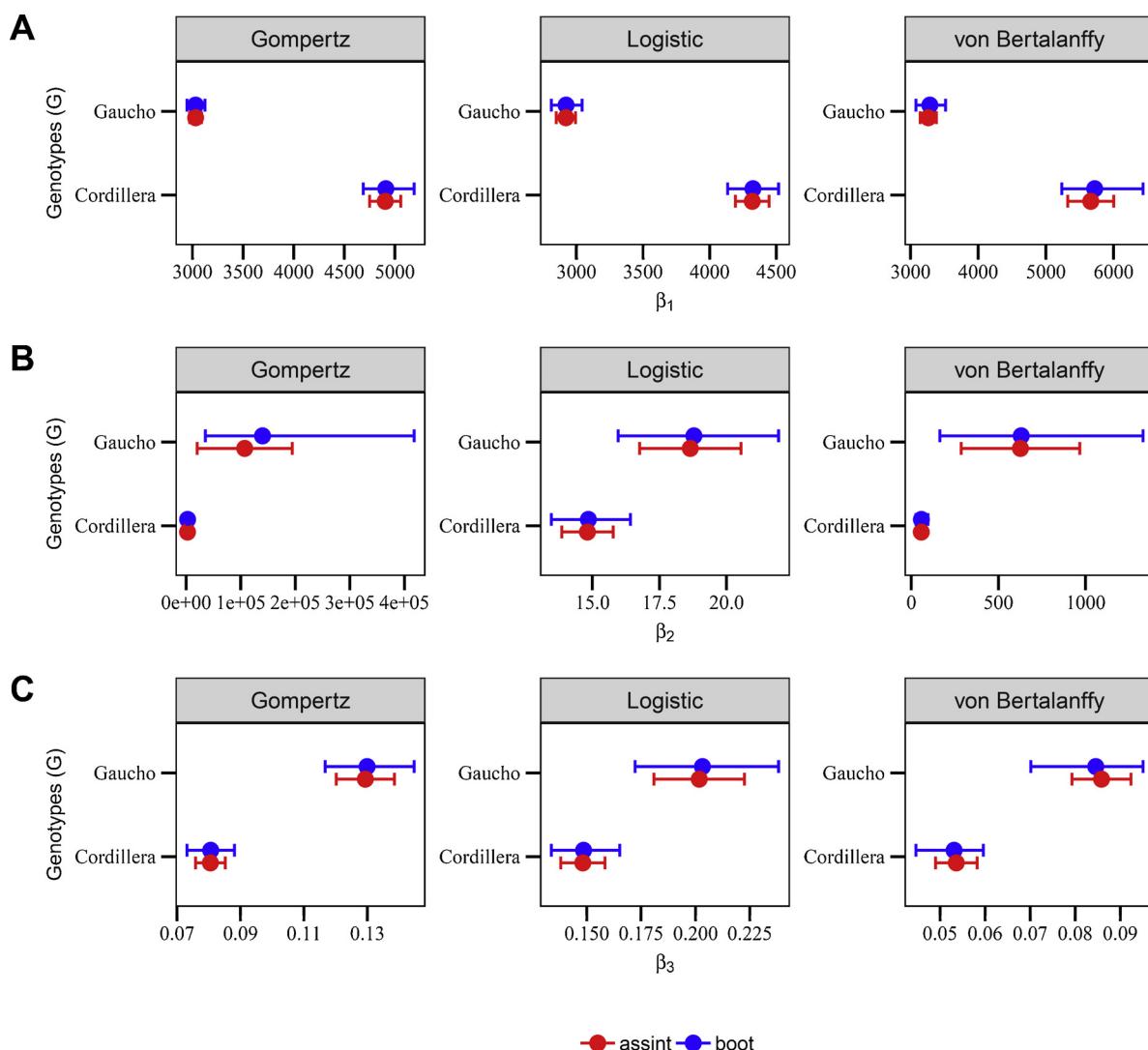


Fig. 7. β_1 (A), β_2 (B) and β_3 (C) estimates and confidence intervals obtained by ordinary least squares (red lines), and Bootstrap-based estimates (blue lines) of the Brody, Gompertz, logistic and von Bertalanffy models fitted to the average mass of fruits per plant on 2016/2017 GS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

approximately 85 DAT in the Cordillera and Ellen genotypes, earlier than in the Santo Clara genotype, which reached the IP at 87 DAT. In the 2016/2017 GS, the Gaúcho genotype reached this point approximately 90 DAT, while the Cordillera genotype reached the IP approximately at 100 DAT (Figs. 14–17).

After the IP, the production rate (velocity) tends to decay (decelerate) to a maximum determined by the maximum deceleration point (MDP). It is observed that, as occurred for the other critical points, the MDP was reached earlier in the Cordillera and Ellen genotypes in the 2016/2017 GS, and by the Gaúcho genotype in the 2016/2017 GS. The asymptotic deceleration point (ADP) indicates the moment of harvest at which increases in production become insignificant. In the 2015/2016 GS, significant increases in yield occurred up to approximately 92 DAT in the Cordillera and Ellen genotypes and up to approximately 97 DAT in the Santa Clara genotype. In the 2016/2017 GS the increments were significant up to approximately 106 DAT in the Gaúcho genotype, and up to approximately 118 DAT in the Cordillera genotype (Figs. 14–17).

4. Discussion

4.1. The ANOVA as a statistical analysis tool in vegetable crops trials

The productive characteristics of the vegetable crops impose a great challenge to the statistical analysis using standard ANOVA procedures. Due to the uneven maturation of fruits, it is observed a high heterogeneity among the harvests, leading a violation of the assumptions of the ANOVA models. The zero-inflated dataset increases the residuals heteroscedasticity and prevents that the data transformation solves this problem.

The linear mixed models can be an alternative in these cases, once allowing attribute differences variances to the harvests through to the variances functions (Lee et al., 2008), according we demonstrate in section results. However, Piepho (2009) argue that the use of the linear mixed models can have several complications for the researcher in relation to the traditional ANOVA such as: i) larger number of alternative variance structures; ii) different standard errors in each pairwise treatment comparison when using models with heterogeneous variance; and iii) the necessity of the approximate degrees of freedom in each treatment comparison. Due to this, the author recommends that an adequate data transformation (that stabilizes the variances) must be

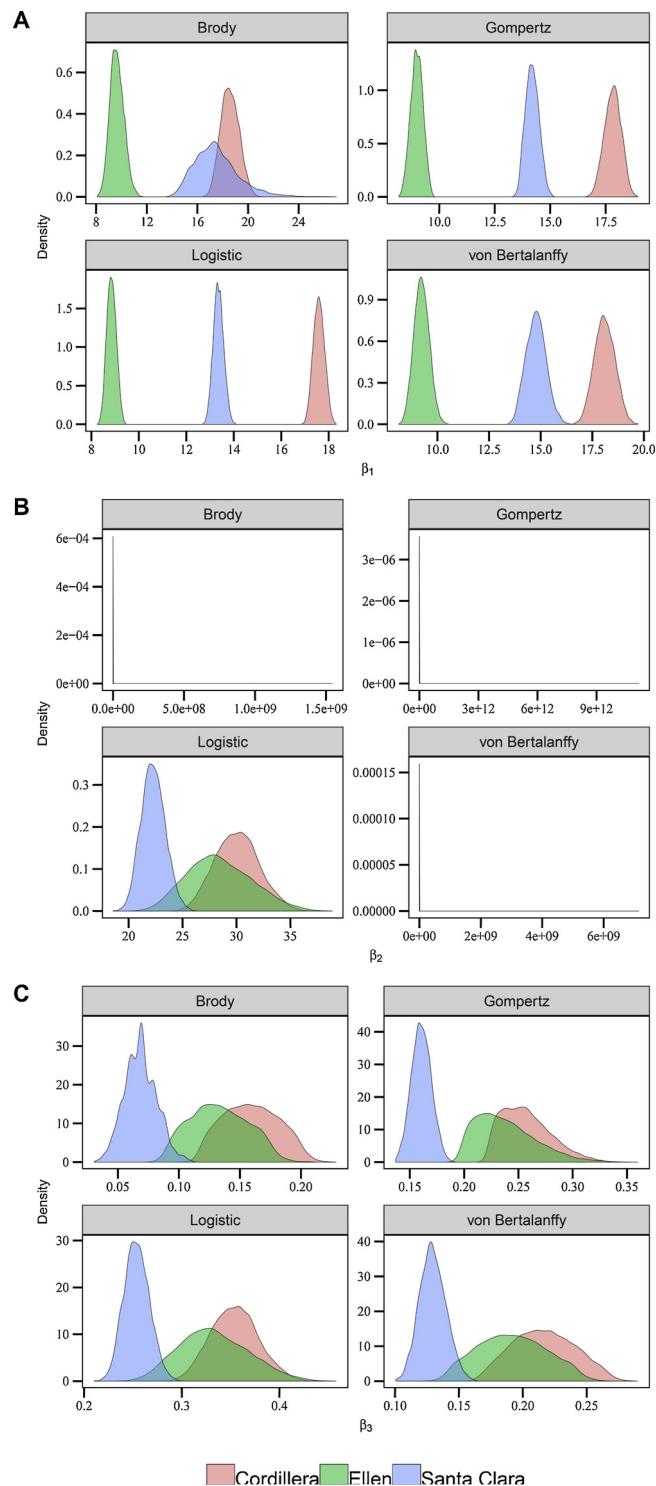


Fig. 8. β_1 (A), β_2 (B) and β_3 (C) bootstrap estimates distribution of the Brody, Gompertz, logistic and von Bertalanffy models fitted to the average number of fruits per plant on 2015/2016 GS.

evaluated after the use the linear mixed models.

How do we could “avoid” the several complications of the linear mixed models when neither the data transformation stabilizes the variances, which was our case? We proposed the use of the nonlinear models. This approach may be performed with the average of the replications, allowing the compare the total production besides increasing the inferences regarding the productive behavior of the multiple harvest crops over time.

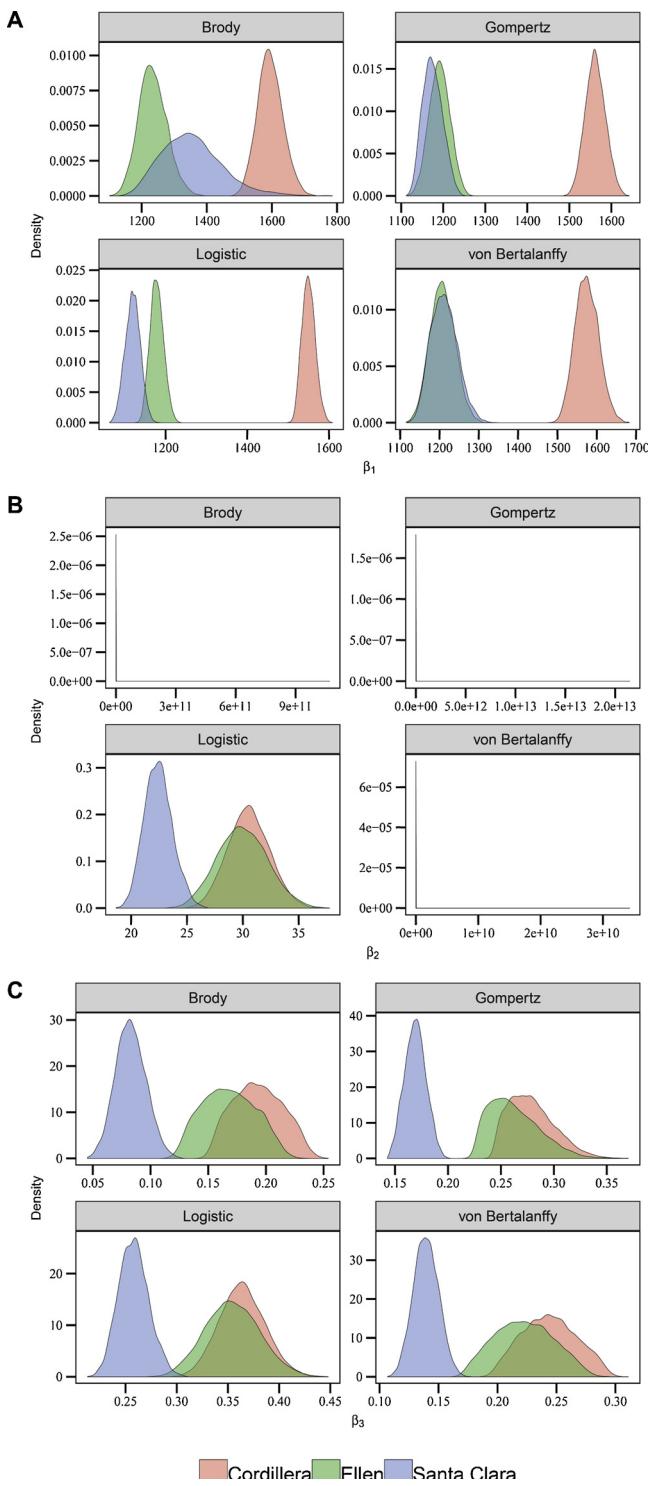


Fig. 9. β_1 (A), β_2 (B) and β_3 (C) bootstrap estimates distribution of the Brody, Gompertz, logistic and von Bertalanffy models fitted to the average mass of fruits per plant on 2015/2016 GS.

4.2. How do we could compare nonlinear models?

As the parameters in nonlinear models are estimated by the OLS method, it is expected that they have the same characteristics of the linear model's parameters. Due to this, we verify that the 10,000 parameter estimates obtained by a bootstrap-approach have normal distribution, which is a characteristic of the linear model's parameters. The same occurs with the confidence intervals obtained by OLS and

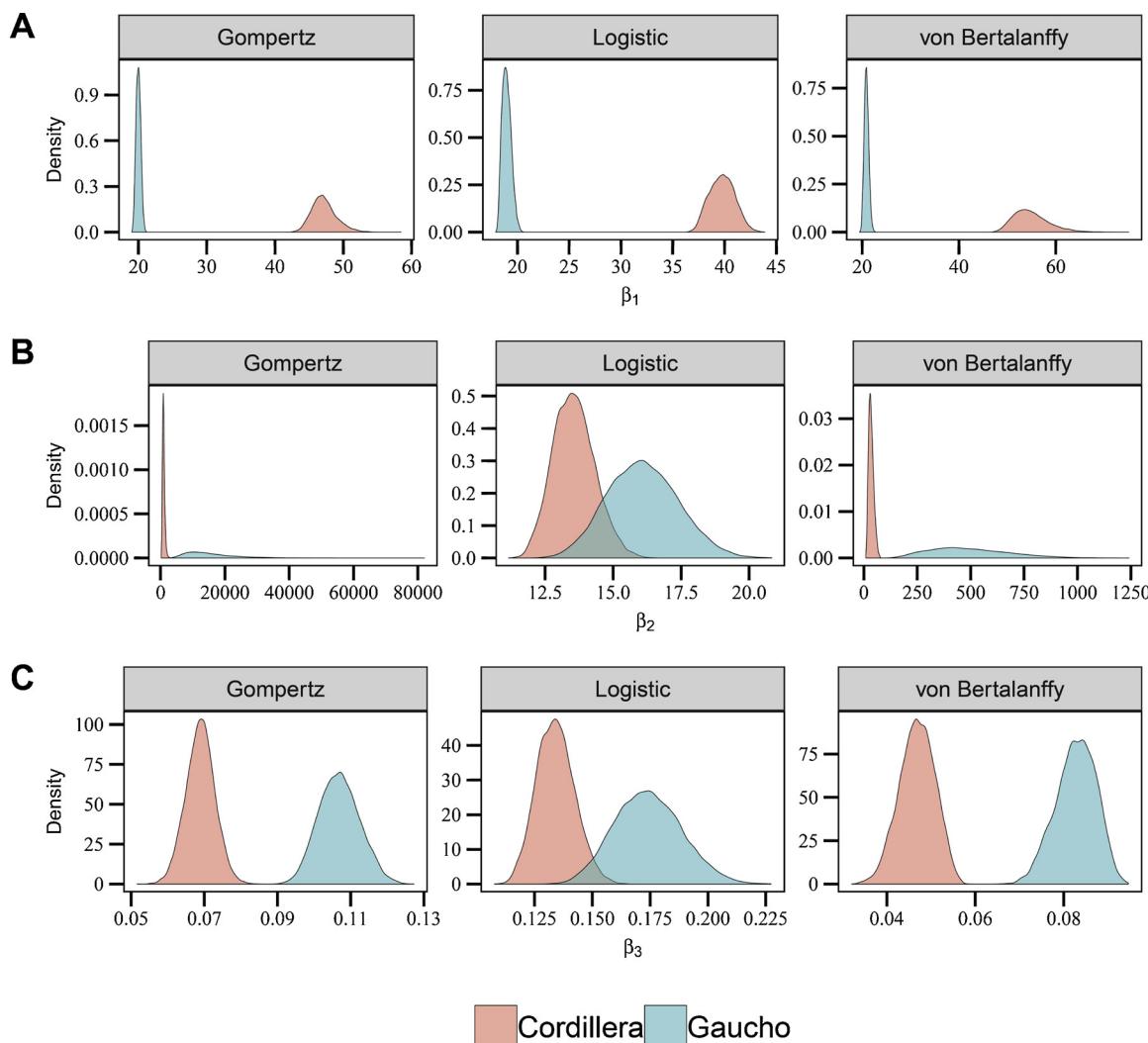


Fig. 10. β_1 (A), β_2 (B) and β_3 (C) bootstrap estimates distribution of the Brody, Gompertz, logistic and von Bertalanffy models fitted to the average number of fruits per plant on 2016/2017 GS.

bootstrap-approach. The differences between the OLS estimates and the mean of the sampling distribution is used to measure the bias of the parameter estimates. By using this approach, we have verified that the Brody, Gompertz e von Bertallanfy have poor linear approximation and high bias in parameter estimates (mainly regarding β_2). Once we have identified highly biased parameter estimates, it should not be logical to make hypothesis tests and discuss the differences between such estimates.

As the parameters are biased, the critical points are estimated wrongly, and the conclusion about precocity and productive concentration are imprecise (remember that the critical points are determinate using the parameter estimates). Besides that, the estimates cannot be compared among the models because they are solutions of distinct equations.

Unfortunately, assessing the bias of parameter estimates has not been used widely in nonlinear modeling. Mostly, the selection of nonlinear models has been performed using goodness-of-fit measures such as AIC, BIC, R^2 , etc (Lúcio et al., 2015, 2016b; Bem et al., 2017a,b). These measures do not asses the bias of the parameters and may lead to selected models that do not describe the biological phenomena correctly. Due to the great power processing of the current computers and the availability of specialized statistical packages to do that, we hope this procedure be popularized and become a standard step in future studies that will use parameters of a nonlinear model to make

inferences regarding any biological process.

We opted by describe the productive behavior of the tomato genotypes from the logistic model because it had parameters close to being unbiased. We also opted by do not compare the results obtained to Brody, Gompertz, and von Bertalanffy models because we determined statistically that these models had biased parameters that could lead to an erroneous interpretation about the productive behavior.

4.3. Biological interpretation of critical points of the models

The asymptote may be used, in a simple and isolated way, to compare the total production of the tomato genotypes, thus replacing the use of ANOVA and post hoc tests. The other parameters and critical points enrich the conclusions of the study, making it possible to make inferences regarding the productive precocity, the concentration of harvests and the productive behavior of the crop over time. However, it is important to emphasize that the interpretation of these parameters should not be done isolated and without a careful analysis of the behavior of the genotypes, as shown below.

The scale parameter (β_2) is directly proportional to the difference between the initial production and the asymptote of the curve and may, therefore, be used to determine the degree of maturation (initial production) of the genotypes at the beginning of the harvest. Comparing two genotypes with the same fruit production rate but different final

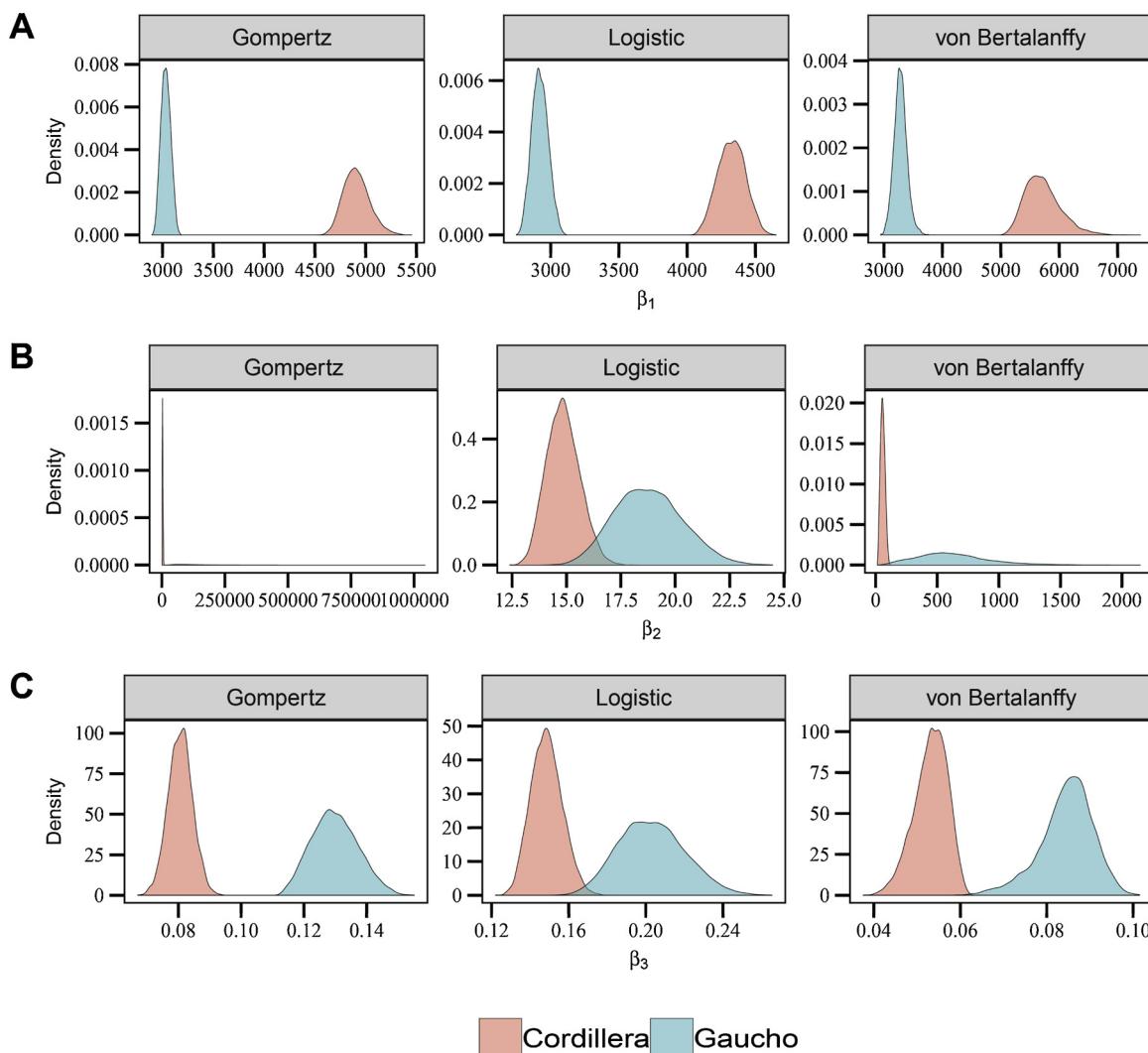


Fig. 11. β_1 (A), β_2 (B) and β_3 (C) bootstrap estimates distribution of the Brody, Gompertz, logistic and von Bertalanffy models fitted to the average mass of fruits per plant on 2016/2017 GS.

production, the one with the highest asymptote (final production) will also present a higher estimate of β_2 if the production of the genotypes at the beginning of the harvests is the same.

It can be seen that even producing a greater mass and number of fruits, the Cordillera genotype presented values of the scale parameter slightly higher than or equal to the other genotypes in the 2015/2016 crop (Fig. 1), indicating an early production (Figs. 3 and 4). The same occurred in the 2016/2017 harvest, when the Cordillera genotype produced a larger number and weight of fruits, but the value of the scale parameter did not differ statistically from the Gaúcho genotype. Although there is no statistically significant difference, the absolute value of the lower scale parameter of the Cordillera genotype indicates an even earlier onset (Fig. 2). It should be noted that since the harvests were started very early in the second experiment, this difference in initial production was not as high (Figs. 5 and 6). Despite the earlier production of the Cordillera genotype in the 2015/2016 GS, the high fruit yield (β_3) of the Ellen genotype (statistically equal to the Cordillera genotype) caused both to reach the IP approximately at the same time.

The Santa Clara genotype reached the IP later because of the lower fruit production rate, showing itself a late genotype (Figs. 3 and 4). In the 2016/2017 GS, the Gaúcho genotype, even showing a late-onset in the production, also reached the IP earlier than the Cordillera genotype due to its higher fruit production rate (Figs. 5 and 6). It is observed that,

even when it is not statistically different, the magnitude of the fruit production rate interferes significantly in the XPI result, and, therefore, the isolated analysis of the parameters and critical points is not recommended to determine the precocity of the genotypes.

The time in which the other critical points (MAP, MDP, and ADP) are reached follow the same behavior of the IP since they are highly influenced by the fruit production rate (Figs. 3–6). However, the abscissa values of these points may be important in determining the productive behavior of the genotypes throughout the harvests. For example, one can use the abscissa of these points to delimit the moment in which the production is concentrated. According to Mischan and Pinho (2014), approximately 21.10% of total production (asymptote) occurs when MAP is reached; 50.00% when the IP is reached; 78.80% when the MDP is reached; and 90.80% when the ADP is reached.

In the 2015/2016 GS, the production of the Cordillera and Ellen genotypes was concentrated in approximately eight days (from 82 to 89 DAT), corresponding to the beginning of the harvest and to the MDP. The remaining 21.20% were produced over 12 days. The production of the Santa Clara genotype included a longer period, concentrating in approximately 10 days, between 92 and 102 DAT, with significant increases in production occurring up to approximately 97 DAT (Figs. 3 and 4). Therefore, the Cordillera and Ellen genotypes are characterized by having a more early and concentrated production than the Santa Clara genotype.

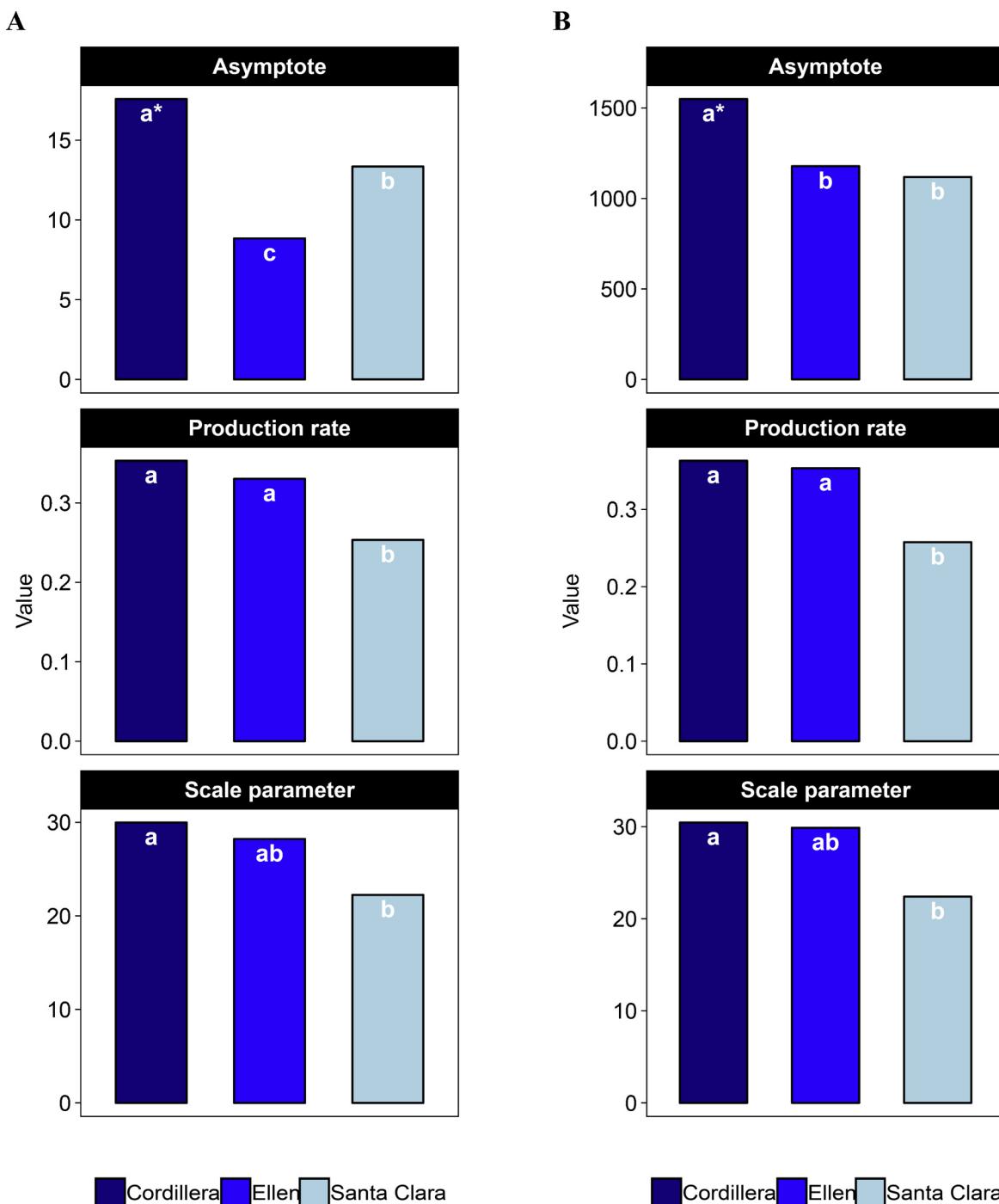


Fig. 12. Parameters of the logistic model for the average number (A) and mass (B) of fruits per plant in three tomato genotypes grown in 2015/2016.

In the 2016/2017 GS, it was observed a slower initial growth, differently than in the previous harvest. MAP was reached in the Cordillera and Gaúcho genotypes at 19 and 14 days, respectively, after the beginning of the harvest. The period between the MAP and the MDP in these genotypes has a similar time period, (18 and 13 days) but they concentrate more than 50% of the production. Finally, the remaining 21.2% is produced in the Cordillera and Gaúcho genotypes at 11 and 21 days, respectively (Figs. 5 and 6). In this experiment, it is clear that the genotype Gaúcho produced earlier and in a more concentrated form than the Cordillera genotype.

The period between MAP and MDP was short and contained a large part of the production of the studied genotypes. Production before and after these points was small and more dispersed over time. Thus,

inferences regarding precocity and crop concentration can also be made using the period between MAP and MDP as a reference. MDP values even indicate until what time production increments are significant. From this moment on, the accumulated production is small (corresponding to less than 10% of the production) and extend for a relatively long period (Figs. 3–6).

5. Conclusions

Alternatives to standard ANOVA procedures are needed when the assumptions of the model are not met. In particular, the data transformation does not improve the assumptions of the analysis of vegetable crops trials due to the excess of zero in the dataset. In this context,

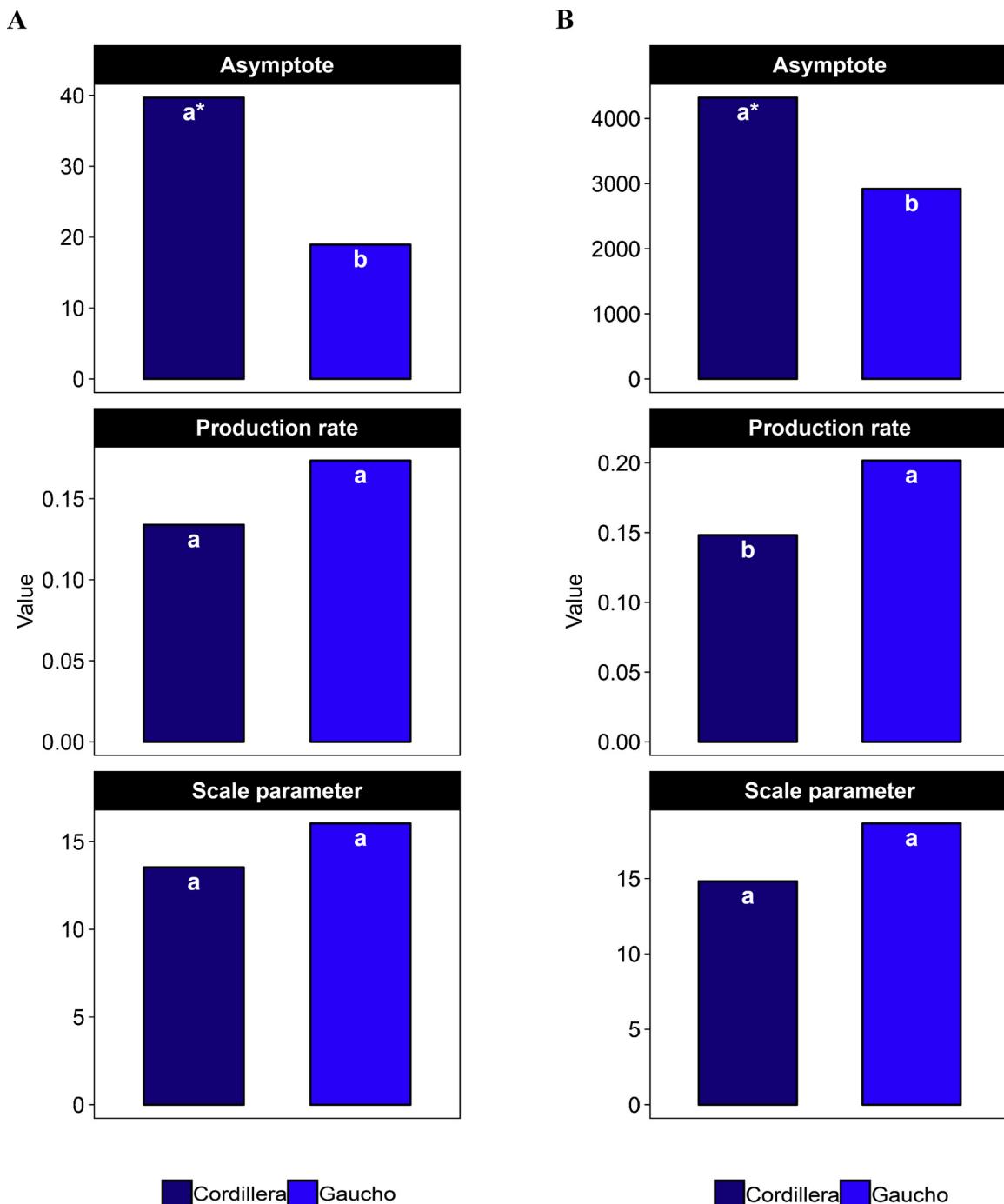


Fig. 13. Parameters of the logistic model for the average number (A) and mass (B) of fruits per plant in three tomato genotypes grown in 2016/2017.

the use of a linear mixed model should be then an interesting alternative. Using mixed model, however, impose a kind of complications to the researcher, such as the large number of alternative variance structures, different standard error for each pairwise comparison and the use of approximation of the degree of freedom of each pairwise comparison. Therefore, we proposed to use the nonlinear models with an alternative the statistical analysis in vegetable crops trials. In this study, the problems that could be caused by heteroscedasticity and dependence of residuals were not perceived. However, if they had been observed, this problem could be solved by the application of adequate statistical methods (such as in linear mixed models). The nonlinear models increase the inferences about the productive behavior of the vegetable crops, and this is an advantage in comparison with ANOVA and linear

mixed models. The comparison of the parameters, through the use of nested models allowed comparing the production and increased the inferential value of the results by determining the initial maturation (through the scale parameter) and the fruit production rate of the genotypes (through parameter β_3). The critical points allowed determining the precocity and concentration of the harvests. Using the parameters and critical points of the model, it was possible to determine not only the final total production but also the maturation of the genotypes at the beginning of the harvest, the productive precocity and the period in which the production is concentrated. The procedures described in this article may be extended for other multiple-harvest crops that have the production described by a sigmoid curve. Thus, it is expected that the inferences of future studies will increase if the

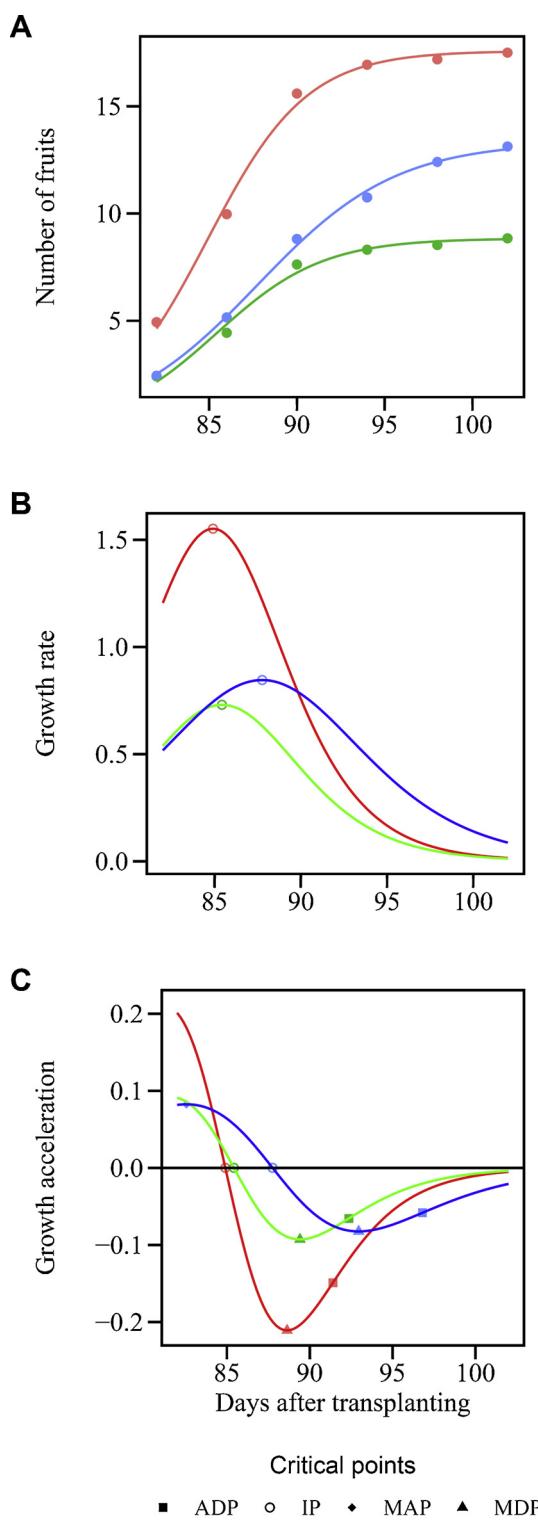


Fig. 14. Logistic model (A), fruit production rate (B) and growth acceleration (C) fitted to the average number of fruits of three tomato genotypes in 2015/2016 GS.

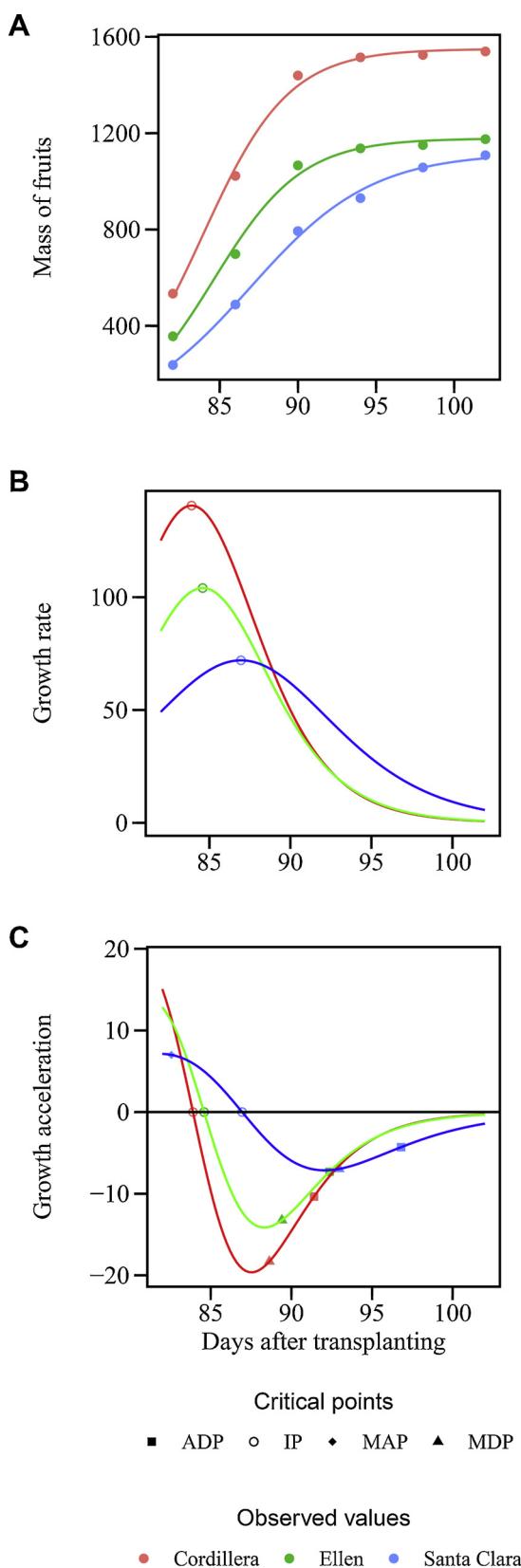


Fig. 15. Logistic model (A), fruit production rate (B) and growth acceleration (C) fitted to the weight of fruits of three tomato genotypes in 2015/2016 GS.

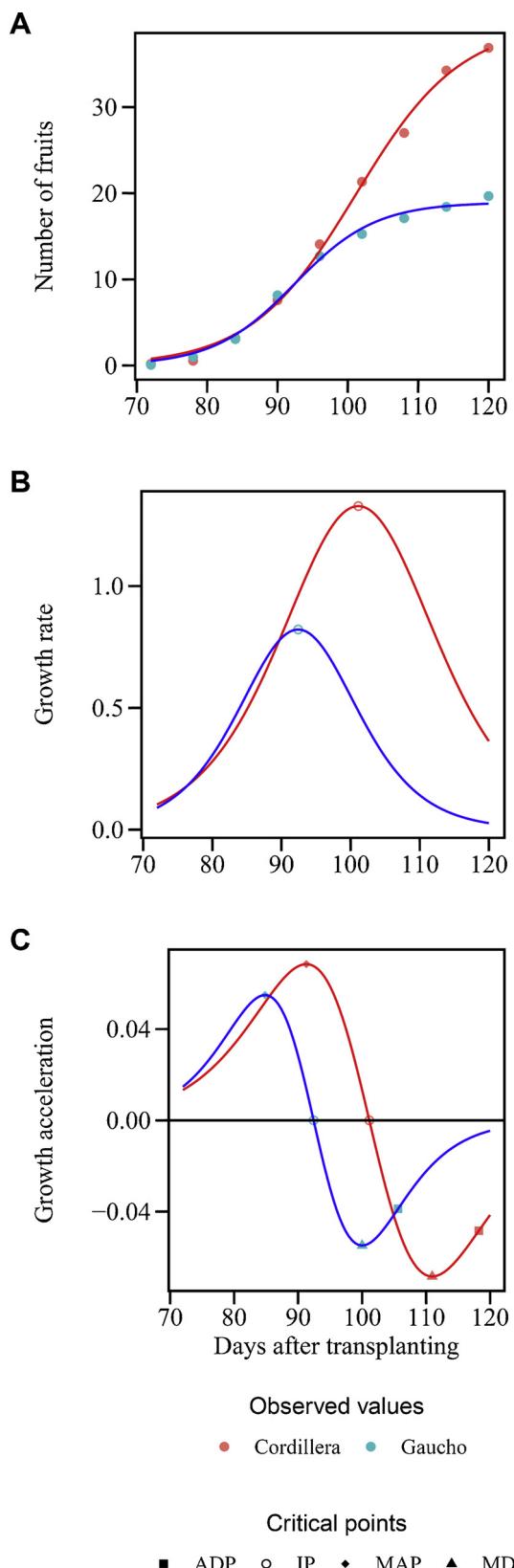


Fig. 16. Logistic model (A), fruit production rate (B) and growth acceleration (C) fitted to the number of fruits of three tomato genotypes in 2016/2017 GS.

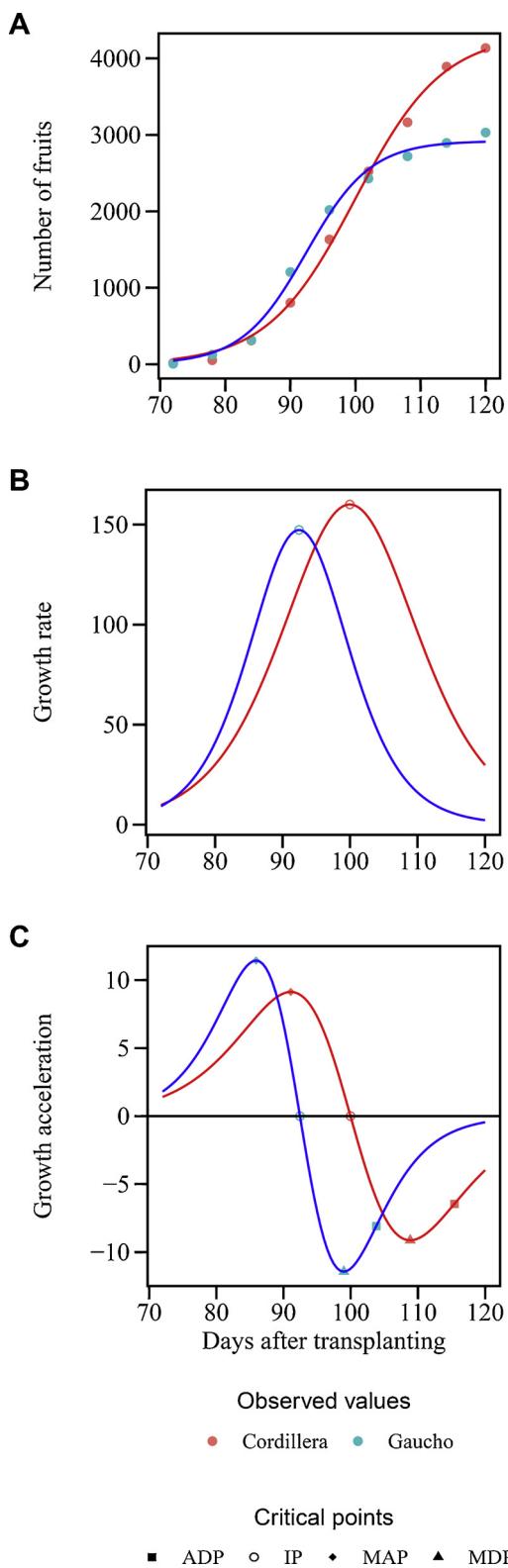


Fig. 17. Logistic model (A), fruit production rate (B) and growth acceleration (C) fitted to the weight of fruits of three tomato genotypes in 2016/2017 GS.

procedures we demonstrated here are used.

Acknowledgments

We thank the Brazilian National Council for Scientific and Technological Development (CNPq) and the Coordination for the

Improvement of Higher Level Personnel (CAPES), for fellowships and grants to the authors.

References

- Archontoulis, S.V., Miguez, F.E., 2015. Nonlinear regression models and applications in agricultural research. *Agron. J.* 107, 786–798. <https://doi.org/10.2134/agronj2012.0506>.
- Bates, D.M., Watts, D.G., 2007. *Nonlinear Regression Analysis and Its Applications*. John Wiley and Sons Inc., New York, pp. 392.
- Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J.P., Delignette-Muller, M.-L., 2015. A toolbox for nonlinear regression in r: the package nlstools. *J. Stat. Softw.* 66, 1–21. <https://doi.org/10.18637/jss.v066.i05>.
- Bem, C.M., Filho, A.C., Facco, G., Schabarum, D.E., Silveira, D.L., Simões, F.M., Uliana, D.B., 2017a. Growth models for morphological traits of sunn hemp. *Semina: Cien. Agrar.* 38, 2933–2944. <https://doi.org/10.5433/1679-0359.2017v38n5p2933>.
- Bem, C.M., Filho, A.C., Chaves, G.G., Kleinpaul, J.A., Pezzini, R.V., Lavezo, A., 2017b. Gompertz and logistic models to the productive traits of sunn hemp. *JAS* 10, 225–238. <https://doi.org/10.5539/jas.v10n1p225>.
- Bi, H., Fox, J.C., Li, Y., Lei, Y., Pang, Y., Bi, H., Fox, J., Li, Y., Lei, Y., Pang, Y., 2012. Evaluation of nonlinear equations for predicting diameter from tree height. *Can. J. For. Res.* 42, 789–806. <https://doi.org/10.1139/X2012-019>.
- Carneiro, A.P.S., Muniz, J.A., Carneiro, P.L.S., Malhado, C.H.M., Filho, R.M., Silva, F.F., 2014. Identidade de modelos não lineares para comparar curvas de crescimento de bovinos da raça Tabapuã. *Pesq. Agrop. Bras.* 49, 57–62. <https://doi.org/10.1590/S0100-204X2014000100008>.
- Ferreira, E.B., Cavalcanti, P.P., Nogueira, D.A., 2018. ExpDes: Experimental Designs. Available at . Accessed in 21 November 2018. <https://CRAN.R-project.org/package=ExpDes>.
- Goldberg, V., Ravagnolo, O., 2015. Description of the growth curve for Angus pasture-fed cows under extensive systems. *J. Anim. Sci.* 93, 4285–4290. <https://doi.org/10.2527/jas.2015-9208>.
- Helidoniotis, F., Haddon, M., Tuck, G., Tarbath, D., 2011. The relative suitability of the von Bertalanffy, Gompertz and inverse logistic models for describing growth in blacklip abalone populations (*Haliotis rubra*) in Tasmania, Australia. *Fish. Res.* 112, 13–21. <https://doi.org/10.1016/j.fishres.2011.08.005>.
- Katsanevakis, S., Maravelias, C.D., 2008. Modelling fish growth: Multi-model inference as a better alternative to a priori using von Bertalanffy equation. *Fish Fish.* 9, 178–187. <https://doi.org/10.1111/j.1467-2979.2008.00279.x>.
- Leduc, D., Goelz, J., 2009. A height-diameter curve for longleaf pine plantations in the gulf coastal plain. *South J. Appl. For.* 33, 164–170. <https://doi.org/10.1093/sjaf/33.4.164>.
- Lee, C.J., O'Donnell, M., O'Neill, M., 2008. Statistical analysis of field trials with changing treatment variance. *Agron. J.* 100, 484–489. <https://doi.org/10.2134/agronj2007.0112>.
- Lúcio, A.D., Sari, B.G., 2017. Planning and implementing experiments and analyzing experimental data in vegetable crops: problems and solutions. *Hortic. Bras.* 35, 316–327. <https://doi.org/10.1590/S0102-053620170302>.
- Lúcio, A.D.C., Couto, M.R.M., Lopes, S.J., Storck, L., 2011. Transformação box-cox em experimentos com pimentão em ambiente protegido. *Hortic. Bras.* 29, 38–42. <https://doi.org/10.1590/S0102-05362011000100007>.
- Lúcio, A.D.C., Nunes, L.F., Rego, F., 2015. Nonlinear models to describe production of fruit in *Cucurbita pepo* and *Capiscum annum*. *Sci. Hortic.* 193, 286–293. <https://doi.org/10.1016/j.scienta.2015.07.021>.
- Lúcio, A.D., Sari, B.G., Pezzini, R.V., Liberalesso, V., Delatorre, F., Faé, M., 2016a. Heterocedasticidade entre fileiras e colheitas de caracteres produtivos de tomate cereja e estimativa do tamanho de parcela. *Hortic. Bras.* 34, 223–230. <https://doi.org/10.1590/S0102-05362016000200012>.
- Lúcio, A.D., Sari, B.G., Rodrigues, M., Beviláqua, L.M., Voss, H.M.G., Copetti, D., Faé, M., 2016b. Modelos não-lineares para a estimativa da produção de tomate do tipo cereja. *Cien. Rural* 46, 233–241. <https://doi.org/10.1590/0103-8478cr20150067>.
- Lúcio, A.D.C., Nunes, L.F., Rego, F., Pasini, M.P.B., 2016c. Relations between zero-inflated variables in trials with horticultural crops. *SJAR* 14, 1–14. <https://doi.org/10.5424/sjar/2016142-8175>.
- Mehätalo, L., De-Miguel, S., Gregoire, T.G., 2015. Modeling height-diameter curves for prediction. *Can. J. For. Res.* 45, 826–837. <https://doi.org/10.1016/10.1139/cjfr-2015-0054>.
- Mendonça, A.R.de, Calegario, N., Silva, G.Fda, Souza, A.Lde, Trugilho, P.F., Carvalho, S.P.Ce, Possato, E.L., 2014. Modelagem da produção de sortimentos em povoamentos de eucalipto. *CERNE* 20, 587–594. <https://doi.org/10.1590/01047760201420041366>.
- Mischan, M.M., Pinho, S.Z., 2014. Modelos não lineares: funções assintóticas de crescimento. *Cultura Acadêmica*, São Paulo pp. 184.
- Mischan, M.M., de Pinho, S.Z., de Carvalho, L.R., 2011. Determination of a point sufficiently close to the asymptote in nonlinear growth functions. *Sci. Agric.* 68, 109–114. <https://doi.org/10.1590/S0103-90162011000100016>.
- Piepho, H.P., 2009. Data transformation in statistical analysis of field trials with changing treatment variance. *Agron. J.* 101, 865–869. <https://doi.org/10.2134/agronj2008.0226x>.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Core Team, R., 2018. . *nlme*: Linear and Nonlinear Mixed Effects Models. Available at . Accessed in 21 November 2018. <https://cran.r-project.org/web/packages/nlme/nlme.pdf>.
- Pinho, S.Zde, Carvalho, L.Rde, Mischan, M.M., Passos, J.R.deS., 2014. Critical points on growth curves in autoregressive and mixed models. *Sci. Agric.* 71, 30–37. <https://doi.org/10.1590/S0103-90162014000100004>.
- R Core Team, 2018. . R: A Language and Environment for Statistical Computing. Available at . Accessed in 21 November 2018. <https://www.R-project.org/>.
- Regazzi, A.J., 2003. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não linear. *Ceres* 50, 9–26.
- Ritz, C., Streibig, J.C., 2008. *Nonlinear Regression With R*. Springer, New York pp. 142.
- ROLAS, 2004. (Rede oficial de laboratórios de análise de solo e de tecido vegetal dos estados do Rio Grande Do Sul e de Santa Catarina). Recomendações de adubações e de calagem para os estados do Rio Grande do Sul e Santa Catarina 224.
- Seber, G.A.F., Wild, C.J., 2003. *Nonlinear Regression*. John Wiley and Sons Inc, New Jersey, pp. 768.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed. Springer, Houston, TX. <https://doi.org/10.1007/978-3-319-24277-4>.
- Wubs, A.M., Ma, Y.T., Heuvelink, E., Hemerik, L., Marcelis, L.F.M., 2012. Model selection for nondestructive quantification of fruit growth in pepper. *J. Am. Soc. Hortic. Sci.* 137, 71–79.