



# Describing tomato plant production using growth models

Bruno Giacomini Sari<sup>a,\*</sup>, Alessandro Dal'Col Lúcio<sup>a</sup>, Cinthya Souza Santana<sup>a</sup>,  
Taciana Villela Savian<sup>b</sup>

<sup>a</sup> Federal University of Santa Maria, Department of Plant Science, Av. Roraima 1000, 97105-900, Santa Maria, RS, Brazil

<sup>b</sup> University of São Paulo, College of Agriculture, Department of Math and Sciences, Av. Pádua Dias 11, 13418-900, Piracicaba, SP, Brazil

## ARTICLE INFO

### Keywords:

*Solanum lycopersicum*  
Non-linear models  
Production growth  
Productive precocity  
Vegetable crops

## ABSTRACT

The aim of this study is to describe the productive behavior of salad tomato genotypes using adjusted growth models. Data were obtained from field experiments performed in 2015/2016 using the Cordillera, Ellen and Santa Clara genotypes, and in 2016/2017 using the Cordillera and Gaúcho genotypes. Five and nine harvests were carried out in 2015/2016 and 2016/2017 respectively, and the variables measured were number and weight of fruit per plant. The Brody, Gompertz, logistic and von Bertalanffy models were adjusted using the accumulated values per plant for each harvest, with the dependent variables being those measured and the independent variable being the number of days after the seedlings had been transplanted. Using the model that best fit the data, the confidence interval of the inflection point was estimated and the similarity of the parameters between the genotypes was measured. The logistic model fit best for both variables. Using the estimates for the biologically interpreted parameters and the inflection point, it was possible to compare the final production of the genotypes and increase the inferences that could be made regarding production over time, differentiating these in terms of productive precocity. In the first experiment, the Cordillera and Ellen genotypes were more premature, while in the second the most premature was the Gaúcho genotype. Therefore, the use of growth models can increase the inferences that can be made in terms of the productive behavior of other crops in multiple harvests.

## 1. Introduction

Some vegetable crops provide more than one harvest over the crop cycle and are known as multiple harvest crops. One of the consequences of this characteristic is that under experimental conditions, production must be measured separately for each harvest, as the variability between the plants will be different for each of these (Lúcio et al., 2015; Lúcio and Sari, 2017). This occurs because the maturation of plant is not uniform. This factor, together with overdispersion and the high number of zero values in the database (Lúcio and Sari, 2017), causes recurring problems in experimental analysis, such as the failure to meet the assumptions of the statistical model in the analysis of variance.

Using harvests (repeated measurements) as a fixed factor can lead to a violation of the independence of residuals in the analysis of variance (ANOVA) since harvests cannot be randomized in the experiment (Fernandez, 2007). In addition, the presence of overdispersion and/or zero-inflated data also leads to violation of assumptions such as normality and homoscedasticity of residuals, making it impossible to

evaluate the data by conventional ANOVA (Fernandez, 2007; Lúcio and Sari, 2017). 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 crops to be carried out based on total production only.

Despite the problems cited above, regression models can be used as a statistical data analysis tool for multiple harvests in vegetable crops (Lúcio et al., 2015, 2016a, 2016b). As the yield values measured for plants in each harvest accumulate, crop production begins to display sigmoidal behavior (Seber and Wild, 2003; Paine et al., 2012), which starts slowly and grows exponentially, before dropping and stabilizing. This type of behavior is typical in non-linear regression models, known as “growth models”. They contain biologically interpreted parameters that quantify the speed of the growth of an organism and determine when this growth tends to stabilize (Seber and Wild, 2003; Mischen et al., 2011).

The growth models are empirical models sufficiently flexible so that

Abbreviations: IP, inflection point; DAT, days after transplant of the seedlings;  $X_{IP}$ , abscissa value for the inflection point;  $Y_{IP}$ , ordinate value of the inflection point; OLS, minimum squares method

\* Corresponding author.

E-mail address: [brunosari@hotmail.com](mailto:brunosari@hotmail.com) (B.G. Sari).

<https://doi.org/10.1016/j.scienta.2018.10.044>

Received 12 July 2017; Received in revised form 25 September 2018; Accepted 18 October 2018

0304-4238/ © 2018 Elsevier B.V. All rights reserved.

fits the data well. The mechanistic models, on the other hand, “that is deduced from the mathematics of the mechanism producing the data” (Seber and Wild, 2003, p. 325). The use of growth models to describe growth is common in agricultural science. The growth models can be used to describe a biological process, such as seed germination (Sousa et al., 2014), fruit growth (Wubs et al., 2012; Muniz et al., 2017v; Fernandes et al., 2017; Ribeiro et al., 2018) or plant growth (Bem et al., 2017v, 2018). However, the nonlinear regression is still little used when statistical analysis in vegetable crops trials, and when it's used, is mostly limited to adjusting growth models to accumulated production data. For example, Lúcio et al. (2015), (2016a) and (2016b) used the logistic and von Bertalanffy models to describe fruit production in zucchini, pepper, cherry tomato, and green beans, and obtained increasing quality adjustments. However, these studies only evaluated the goodness-of-fit the models to the data, and do not explore the physical interpretation of the parameters and critical points.

The empirical models, such as growth models, fit it a sigmoidal pattern data (Seber and Wild, 2003). The goodness-of-fit of these models is related to the parameters estimate of the models, so it can be used to describe the growth of the production. Besides that, the parameter estimates can be used to compare the behavior of production among treatments. The advantages of these models, in comparison polynomial models, are the physical interpretation of the parameters of the growth models (Seber and Wild, 2003). According to Seber and Wild (2003), Ritz and Streibig (2008) and Archontoulis and Miguez (2015), dichotomous variables can be used to verify whether a set of models has the same parameter values via the F test, which compares parameters between models. Puiatti et al. (2013) used model identity in this way to verify the similarity of the development of bulbs in garlic genotypes.

In addition to parameters involving biological interpretation, growth models also indicate critical points on the curve that can be extremely important explanatory variables in terms of productive behavior. One of these is the inflection point, whose ordinate determines an organism's maximum speed of growth (Seber and Wild, 2003; Mischán et al., 2011). Therefore, the abscissa of the inflection point can be determined in order to predict the harvest time at which the plant produces the most fruit. The inflection point indicates the plants' maximum rate of production and the time at which this occurs. This information can be used to determine genotype early yield, and plan future tomato crops.

Determine the precocity and the concentration of the production is important to identify the potential for early money returns to the farmer (Khanizadeh and Fanous, 1992). Khanizadeh and Fanous (1992) developed an early ripening and concentrated cropping index to assess the precocity and the concentration of production. However, the rate of the production and the moment at the maximum production occur do not are determinate by this mathematical indices.

Thus, besides describe the productive behavior of tomato plant, the growth models (through its parameters and critical points) can be used to determine the productive precocity exactly. The productive precocity can be defined exactly like the moment of occurrence of the maximum rate of production using the inflection point of the growth curves. Thus, independent of the productive behavior of any multiple harvest crops, the nonlinear models were given a mathematical solution to a characteristic that is determined subjectively (in most of the case). Besides that, a unique values (parameter value) can be associated to the occurrence of the maximum rate of production and, consequently, to the productive precocity.

One of the models that describe tomato production is TOMGRO (Jones et al., 1991). However, it is a mechanistic model designed to predict yield. In contrast, growth models are empirical and provide information on the increase in fruit production as time progresses. Growth models are capable of describing the sigmoidal behavior of fruit production, and the biological interpretations of parameters and critical points can be used to obtain explanatory variables for the productive

behavior of the tomato over time.

The sigmoidal behavior of production over time allows growth models to be adjusted, and can be used as an important statistical tool, increasing the number of inferences regarding productive behavior, based on an interpretation of model parameters, and overcoming the limitations of the ANOVA (harvests cannot be randomized in the experiment as a factor, presence of overdispersion and/or zero-inflated data in vegetable crops trials). The aim of this study is to describe the productive behavior of tomato genotypes (*Solanum lycopersicum*) using adjusted growth models and determine the productive precocity through the parameters and critical points of interpretation.

## 2. Materials and methods

### 2.1. Site description and experimental design

Data were taken from two field experiments performed at the Department of Plant Science at the Federal University of Santa Maria (UFSM), Rio Grande do Sul state, Brazil (29°43'S, 53°43'O, altitude 95 m) during 2015/2016 and 2016/2017. The Cordillera, Ellen and Santa Clara hybrids were used for the 2015/2016 experiment and the Cordillera and Gaúcho hybrids were used for the 2016/2017 experiment. The seedlings were transplanted on October 17, 2015, in the first experiment, and on September 10, 2016, in the second experiment. The spacing of rows between the seedlings was 1.00 and 0.30 m for the 2015/2016 experiment and 1.00 and 0.50 m for the 2016/2017 experiment. The plants were staked using polypropylene cord and a single stake in both experiments. Fertilizing was based on chemical analysis of the soil and followed recommendations by the Official Network of Soil and Plant Tissue Analysis Laboratories for the States of Rio Grande do Sul and Santa Catarina (ROLAS, 2004). Pest and disease management followed the technical recommendations for the crop. Climatic data was collected by a weather station situated 200 m from the experiment location.

Harvests were carried out 82, 86, 90, 94, 98 and 102 days after transplant of the seedlings (DAT) for the 2015/2016 experiment and 72, 78, 84, 90, 96, 102, 108, 114 and 120 DAT for the 2016/2017 experiment. Fruits were collected when they changed from green to red in color and were then packed in bags and taken to the laboratory, where they were counted and weighed using digital scales with milligram precision.

### 2.2. Adjustment and selection of models

The measured values for weight and quantity of fruit per plant for each harvest were accumulated for each genotype, after which the following growth models were adjusted: a) Brody:  $Y_i = \beta_1 - (\beta_2 e^{(-\beta_3 t_i)}) + \varepsilon_i$ ; b) Gompertz:  $Y_i = \beta_1 e^{(-\beta_2 e^{(-\beta_3 t_i)})} + \varepsilon_i$ ; c) Logistic:  $Y_i = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 t_i)}} + \varepsilon_i$ ; and d) von Bertalanffy:  $Y_i = \beta_1 (1 - \beta_2 e^{(\beta_3 t_i)})^3 + \varepsilon_i$ ; in which:  $Y_i$  is the value of the accumulated variables (number and weight of fruit) over time (dependent variable);  $t_i$  is the time interval between transplant and the harvest (independent variable);  $\beta_1$  is the parameter representing the asymptotic value;  $\beta_2$  reflects the distance between the initial and the asymptotic value; and  $\beta_3$  is associated with the rate of growth.

Parameters were estimated using the ordinary least-squares method, using a Gauss-Newton approach and solving ordinary least square at each iteration. Following this, the Shapiro-Wilk and Breusch-Pagan tests were applied to verify the normality and homogeneity of the residuals respectively. Finally, the independence of the residuals can be tested by the lag plot of the  $n$ th residual versus  $(n-1)$ th residual, according to Bates and Watts (2007). The quality of the fit of each adjusted model for each genotype was determined using the coefficient of determination ( $R^2$ ) and through verification of the nonlinearity of the model, using the Bates and Watts curvature method (Bates and Watts,

2007). The model that presented the highest  $R^2$  value (close to 1) and low nonlinearity (less than 1) was selected.

### 2.3. Determination of the inflection point

The inflection point of the model that best fit the weight and number of fruit per plant data was calculated for each genotype. The inflection point (IP) was obtained by setting the second order derivative of the model equal to zero (Mischan et al., 2011). After this, the confidence interval for the abscissa of the inflection point was calculated using the following equation (Pinho et al., 2014):

$$IC(X_{IP}; 95\%) = X_{PI} \pm S(IP) \times t_{(0.05; GL_{Error})} \quad (1)$$

Where  $X_{IP}$  is the abscissa of the inflection point,  $t_{(0.05; GL_{Error})}$  is the value of the Student's t-distribution quantile with  $\alpha = 5\%$  and  $GL_{Error}$  degrees of freedom, and  $S(IP)$  is the standard deviation of the inflection point calculated for:

$$S(IP) = \sqrt{1/\hat{\beta}_3^2 \{fbb + (\hat{\beta}_2/\hat{\beta}_3)^2 fcc\} - [2(\hat{\beta}_2/\hat{\beta}_3) fbc]} \quad (2)$$

Where  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  and  $\hat{\beta}_3$  are the estimates for the model's parameters and fbb, fcc and fbc correspond to  $\hat{V}(\hat{\beta}_2)$ ,  $\hat{V}(\hat{\beta}_3)$  and  $\hat{Cov}(\hat{\beta}_2, \hat{\beta}_3)$  respectively, these being the estimates for the variance (and covariance) for these parameters.

### 2.4. Parameters comparisons

To verify the similarity between the parameters of the model that best fit the weight and number of fruit data for the genotypes evaluated, model identity tests were performed using a dummy variable ( $D_j$ ):

$$D_j = \begin{cases} D_j, & \text{if the observation pertence to genotype } k; \\ 0, & \text{in any other case.} \end{cases}$$

Where  $j$  is the genotype evaluated. The purpose of this test was to compare a complete model ( $\omega$ ), presenting different parameters for each genotype, with a reduced model ( $\Omega$ ), presenting similar parameters between genotypes. Using the logistic model as an example, the complete model would be as follows:

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

Where  $t_i$  represents the number of day after the transplant of the  $n$ th harvest;  $j = 1, 2, \dots$ , and  $k$  indicates that the observation was made for genotype  $k$ . Note that in this case, there is one parameter for each genotype ( $j$ ). A reduced model would be as follows:

$$Y_{ij} = \sum_{j=1}^k \left[ \frac{\beta_1}{1 + e^{(\beta_2 j - \beta_3 j t_i)}} \right] + \varepsilon_{ij} \quad (4)$$

Not that in (4), the parameter  $\beta_1$  is the same for all  $k$  genotypes. The null hypothesis would be  $H_0: \beta_{11} = \beta_{12} = \dots = \beta_{1k}$ . Using the methodology described above, it was verified whether the null hypotheses were rejected or not for both variables, considering a significance level of 5%:

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

A comparison of the complete and reduced models was performed using the F test (Ritz and Streibig, 2008). The value of  $F_{calc}$  was calculated as follows:

$$F_{calc} = \{[SQ_{Error}(\Omega) - SQ_{Error}(\omega)]/[GL_{Error}(\Omega) - GL_{Error}(\omega)]\}/QM_{Error}(\omega) \quad (5)$$

where  $SQ_{Error}(\Omega)$  is the sum of squares of the errors of the reduced model,  $SQ_{Error}(\omega)$  is the sum of squares of the errors of the complete model and  $QM_{Error}(\omega)$  is the mean squared error of the complete model. The statistic  $F_{calc}$  has a distribution of  $F_{[v, GL_{Error}(\omega)]}$  over  $H_0$ , where  $v$  is the difference between  $GL_{Error}$  for the complete and the reduced model.

The parameters effect on the selected model was represented graphically, from simulated data, to facilitate the interpretation of the results. For this, graphs were constructed varying only one of the model parameters, keeping the others constant. The productive behavior of the genotypes over time was represented in graph form in order to facilitate interpretation of the results.

All analyses were performed using R software (R Core Team, 2014). The function *nls* (R Core Team, 2014) were used to fit the models. The functions *shapiro.test* (R Core Team, 2014) and *bptest* (*lmtest* package, Zeileis and Hothorn, 2002) were used to test the normality and homoscedasticity of the residuals, respectively. Finally, the function *acf* were used to compute the estimates of the autocorrelation function (R Core Team, 2014).

## 3. Results and discussion

### 3.1. Model selection

All the models fit the weight and number of fruit per plant data for the experiment conducted in 2015/2016. For the experiment conducted in 2016/2017, the Brody model could not be adjusted to the two genotypes evaluated. The assumptions of normality, homoscedasticity (Tables 1 and 2), and independence of the residuals were met in most of the cases (excepted to the genotypes Santa Clara and Ellen, in first season, see in Supplemental material), indicating that the estimation of the parameters using the ordinary least-squares method was adequate.

There was a difference in productive behavior between the genotypes for the two years. The first point to be highlighted is the time that harvesting began, which occurred later for the first experiment than for the second. In 2015/2016, the degree of ripening for the genotypes was higher when harvesting began. This reflects the absence of slower initial growth and the possibility of the Brody model fitting the data from the first experiment. This model represents data for which the growth

**Table 1**

P-values for the Shapiro-Wilk (SW) and Breusch-Pagan (BP) tests applied to the residuals, the coefficient of determination ( $R^2$ ) and the parametric nonlinearity ( $c^0$ ) of the Brody, Gompertz, logistic and von Bertalanffy models in three tomato genotypes, for the variables of weight (in grams) and number of fruits per plant, in 2015/2016.

Genotypes	Nº of fruits				Weight of fruit			
	SW	BP	$R^2$	$c^0$	SW	BP	$R^2$	$c^0$
Brody								
Cordillera	0.95	0.09	0.96	444.47 <sup>1</sup>	0.99	0.08	0.97	462.5
Ellen	0.83	0.11	0.94	485.76	0.95	0.08	0.96	467.64
Santa Clara	0.06	0.5	0.98	213.26	0.48	0.33	0.98	200.06
Gompertz								
Cordillera	0.91	0.07	0.98	292.06	0.87	0.09	0.98	301.89
Ellen	0.98	0.09	0.97	346.58	0.99	0.08	0.98	316.14
Santa Clara	0.97	0.11	0.99	95.29	0.89	0.08	0.99	103.98
Logistic								
Cordillera	0.98	0.07	0.99	0.58	0.98	0.11	0.99	0.48
Ellen	0.98	0.08	0.98	0.97	0.95	0.09	0.99	0.63
Santa Clara	0.91	0.12	0.99	0.73	0.87	0.17	0.99	0.76
von Bertalanffy								
Cordillera	0.83	0.08	0.98	323.96	0.88	0.09	0.98	340.26
Ellen	0.94	0.09	0.97	371.41	0.94	0.08	0.97	348.42
Santa Clara	0.65	0.22	0.99	119.67	0.93	0.14	0.99	121.58

<sup>1</sup> Values obtained using the equation  $c^0 \times \sqrt{F_{(5\%, 3, 3)}}$ .

**Table 2**

P-values for the Shapiro-Wilk (SW) and Breusch-Pagan (BP) tests applied to the residuals, the coefficient of determination ( $R^2$ ) and the parametric nonlinearity ( $c^6$ ) of the Brody, Gompertz, logistic and von Bertalanffy models in three tomato genotypes, for the variables of weight (in grams) and number of fruits per plant, in 2016/2017.

Genotype	N° of fruits				Weight of fruit			
	SW	BP	$R^2$	$c^6$	SW	BP	$R^2$	$c^6$
Brody								
Cordillera	.1	–	–	–	–	–	–	–
Gaucha	–	–	–	–	–	–	–	–
Gompertz								
Cordillera	0.34	0.1	0.99	18.21 <sup>2</sup>	0.89	0.12	0.99	17.57
Gaucha	0.45	0.67	0.99	19.8	0.74	0.51	0.99	29.45
Logistic								
Cordillera	0.45	0.46	0.99	0.81	0.79	0.37	0.99	0.6
Gaucha	0.08	0.64	0.99	0.56	0.67	0.87	0.99	0.5
von Bertalanffy								
Cordillera	0.49	0.14	0.99	23.24	0.75	0.16	0.99	15.18
Gaucha	0.94	0.2	0.99	15.31	0.31	0.21	0.99	11.94

<sup>1</sup> The Brody method did not fit the data.

<sup>2</sup> Values obtained using the equation  $c^6 \propto \sqrt{F_{(5\%, 6, 3)}}$ .

rate tends to decrease from the start, and this behavior seems to be more plausible for the first rather than the second experiment (Figs. 1 and 2).

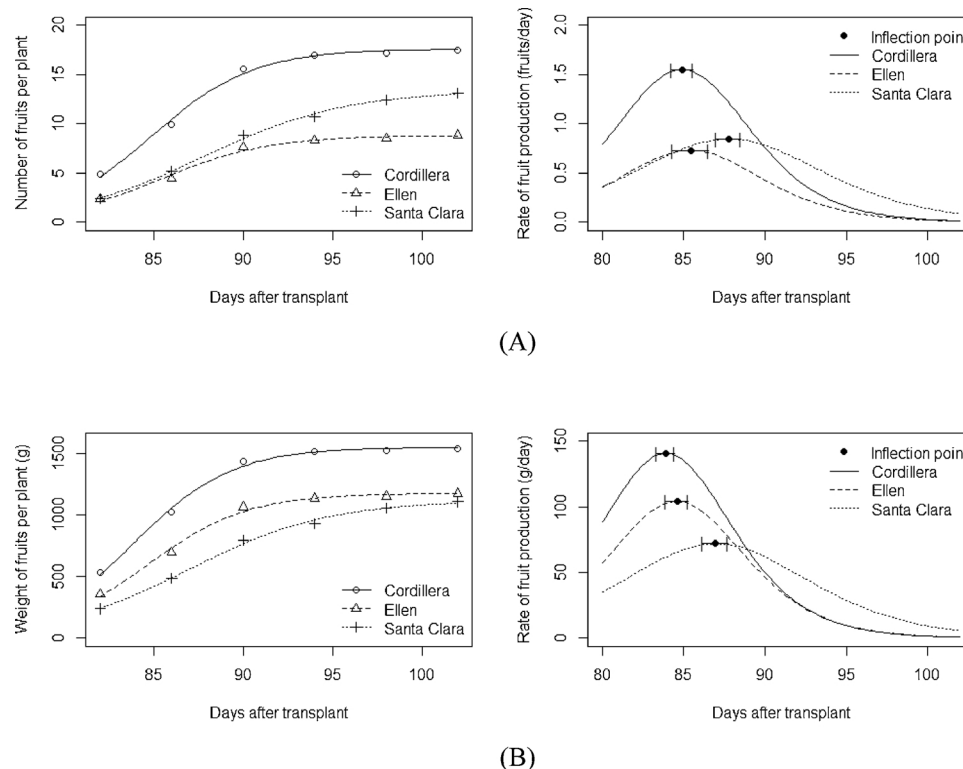
Despite of the models have met the assumptions of normality, homogeneity, and independence of the residuals in most of the cases, and all of them demonstrating higher goodness-of-fit (high  $R^2$ ), the logistic model presented the lowest parametric nonlinearity (Tables 1 and 2). Therefore, independent of the difference in behavior between the experiments mentioned above, the logistic model is the most adequate in describing the production data on the genotypes evaluated. As the estimates for the non-linear model parameters are obtained via a linear approach using the least-squares method, the behavior of the

non-linear model should be close to that of the linear model. In this case, the estimators take on the properties of linear model parameters, are more precise and estimates for biologically inferred parameters are more reliable (Seber and Wild, 2003; Bates and Watts, 2007; Ritz and Streibig, 2008). The value of the curvature that determines if a non-linear model has parametric low linearity or not is controversial, but values below 1 are acceptable (Seber and Wild, 2003).

All of the adjusted models presented a high degree of adjustment (high  $R^2$ ) and lower intrinsic nonlinearity, however the logistic method presented the lowest curvature due to the effect of the parameters (low parametric nonlinearity measures) (Seber and Wild, 2003; Bates and Watts, 2007), and was selected for this reason (Tables 1 and 2). Since the parameters are to be used to describe and compare the productive behaviors of the genotypes, the predictive capability of a non-linear model (determined by high  $R^2$ ) should also include close to begin unbiased parameters estimate (in line with the real observations). Therefore selecting growth models purely on the basis of their predictive capabilities can lead to erroneous conclusions if the parameter values are not open to biological interpretation.

### 3.2. Limitation in the interpretation of the parameters and critical points

The growth models can be used to describe a biological process through the use of the parameters and the critical points of the models. However, the most of the manuscript only identify the model with best goodness-of-fit (Bem et al., 2017v; Fernandes et al., 2017; Ribeiro et al., 2018; Bem et al., 2018). Besides that, the interpretation of the parameters is controversial and many authors do not use them in your analysis. The  $\beta_2$  is an example, since many authors describe it as a parameter without biological interpretation (Lúcio et al., 2015, 2016a; Lúcio et al., 2016b; Bem et al., 2017v, 2018), and this limits the inferences about the biological process. Finally, the use of the critical points and the derivative of the growth model allow describe the sigmoidal behavior of the variables with more details (Godoy et al., 2008),



**Fig. 1.** Observed value (points), adjusted logistic model (lines), rate of fruit production and 95% confidence interval for the inflection point of three tomato genotypes for the variables number (A) and weight of fruit per plant (B) in 2015/2016.



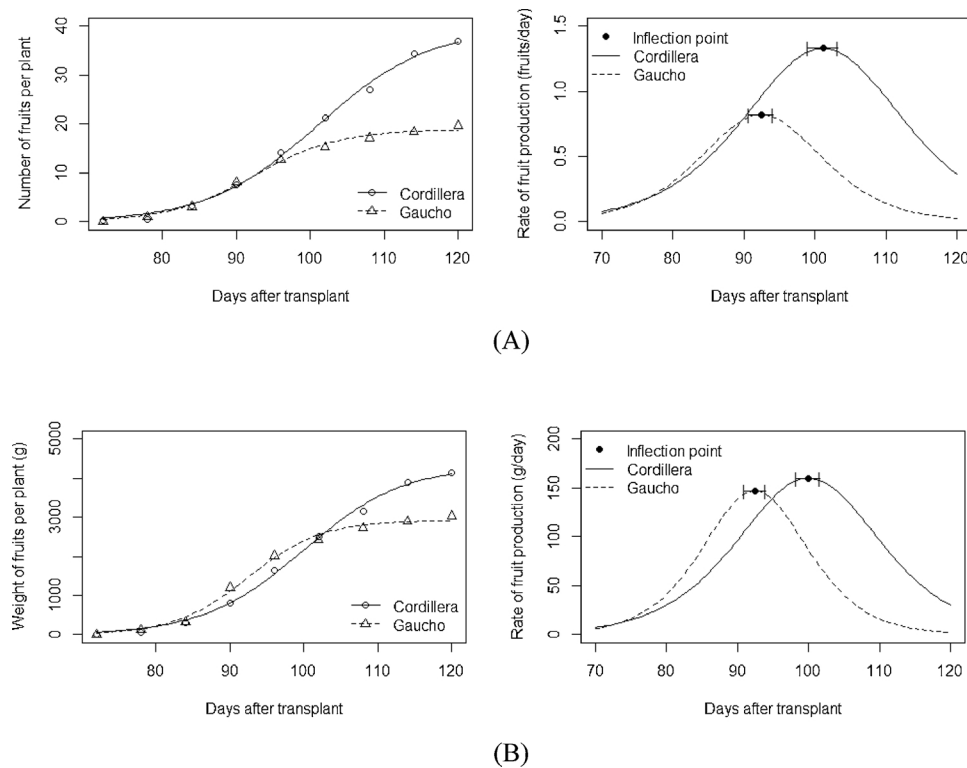


Fig. 2. Observed values (points), adjusted logistic model (lines), the rate of fruit production and 95% confidence interval for the inflection point of three tomato genotypes for the variables number (A) and weight of fruit per plant (B) in 2016/2017.

but it is constantly ignored in many manuscripts.

Another important aspect in studies with nonlinear models is the model selection. The selection criteria in most of the manuscript are performed by the goodness-of-fit measures (Wubs et al., 2012; Lúcio et al., 2015, 2016a, 2016b; Bem et al., 2017v, 2018). These statistics do not guarantee that the parameters are close to being unbiased. Parameters are close to being unbiased is very important why a feature that it describes the biological phenomena with coherence. Due to this, the selection of the models can be performed by the nonlinearity measures so that the result obtained is consistent with reality.

### 3.3. Interpretation of parameters

The estimated values for the asymptote  $\beta_1$  of the model can be considered to be the estimates for total production for the genotypes evaluated. For the 2015/2016 crop, the Cordillera genotype produced an average of 17 fruits per plant, with a total weight of 1549.72 g, differing statistically from the other genotypes. The average weight of fruit per plant for the Ellen and Santa Clara genotypes of around 1100 g did not differ statistically. However, the average number of fruits per plant was higher for the Santa Clara genotype (approximately 13 fruits per plant) compared to the Ellen genotype (approximately 8 fruits per plant), demonstrating that the average weight of the fruit of the latter genotype is higher than the former. For the 2016/2017 crop, the Cordillera genotype produced an average of 39 fruits per plant, with a total weight of 4320.10 g, differing statistically from the Gaucho genotype, which produced an average of 19 fruits per plant and had a total weight of 2921.10 g (Tables 3 and 4).

The lower production in 2015/2016 occurred due to excessive rainfall during the fructification stage of the crop, while in 2016/2017, the rains occurred mainly during the vegetation period (Fig. 3). This together with the fact that the plant population was higher for the 2015/2016 experiment compared to the 2016/2017 experiment explains this large difference in terms of weight and number of fruits per plant for the two years. Note that average yield per hectare was equal to

or higher than that observed in the state of Rio Grande do Sul, Brazil (IBGE, 2017).

The parameter  $\beta_1$  are related to the total production of the genotypes, as already reported. The change in the estimates of  $\beta_1$  does not shift the curve in relation to the abscissa axis (Fig. 4A) and also does not affect the moment of PI occurrence (Fig. 5A). The lower estimate of  $\beta_2$  and a higher estimate of  $\beta_3$  are related to the early production of the genotypes. The relationship of the parameters  $\beta_2$  and  $\beta_3$  with the early production is given by the shift (in relation to abscissa) and greater slope of the growth curve (Fig. 4B and C). It is also possible to verify the early production by the moment of PI occurrence ( $X_{IP}$  values) (Fig. 5B and C). The  $X_{IP}$  is the time at which of fruit production rate is maximal. As the logistic model is symmetrical (Seber and Wild, 2003),  $X_{IP}$  is also the time when 50% of genotype production has already been harvested.

When the estimate of  $\beta_2$  is smaller (keeping the other parameters constant) the growth curve shifts to the left, showing that the production increase tends to begin earlier (Fig. 4B). Similarly, the IP occurs earlier (lower  $X_{IP}$ ) when the estimate of  $\beta_2$  is smaller (Fig. 5B). The same behavior is observed when the estimative of the  $\beta_3$  is an increase. The growth curve shifts to the left due to the higher slope (higher fruit production rate), and that this indicates the increase in production is higher and occurs earlier (Fig. 4C). Similarly, the IP occurs earlier (lower  $X_{IP}$ ) when the estimate of  $\beta_3$  are higher (Fig. 5C). The parameter  $\beta_3$  associate the early production with the higher production rate, whereas the parameter  $\beta_2$  does not (Fig. 5B and C).

The parameters  $\beta_2$  and  $\beta_3$  should be analyzed together. As reported above, a lower estimate of  $\beta_2$  and a higher estimate of  $\beta_3$  indicate earlier production. When both estimates of  $\beta_2$  and  $\beta_3$  are high, as in case of this study (Table 3), it was concluded that the fruit production is early. Therefore, the higher fruit production rate ( $\beta_3$ ) is the component of the model that has a greater influence on earlier production in this case (Figs. 1 and 2).

For the 2015/2016 crop, the comparison of Cordillera and Santa Clara genotypes shows well the relationship between  $\beta_2$  and  $\beta_3$  parameters. For the Cordillera genotype, the estimates of  $\beta_2$  and  $\beta_3$  were

**Table 3**

Parameters of the logistic growth models and estimated inflection point coordinates for tomato genotypes grown in 2015/2016 and 2016/2017, for the variables of number and weight of fruit (in grams) per plant.

Genotype	N° of fruits					Weight of fruit				
	$\beta_1$	$\beta_2$	$\beta_3$	$X_{IP}^1$	$Y_{IP}^2$	$\beta_1$	$\beta_2$	$\beta_3$	$X_{IP}$	$Y_{IP}$
2015/2016										
Cordillera	17.58 <sup>*</sup>	29.98 <sup>*</sup>	0.35 <sup>*</sup>	84.91	8.79	1549.72 <sup>*</sup>	30.45 <sup>*</sup>	0.36 <sup>*</sup>	83.90	774.86
Ellen	8.84 <sup>*</sup>	28.22 <sup>*</sup>	0.33 <sup>*</sup>	85.42	4.42	1178.19 <sup>*</sup>	29.88 <sup>*</sup>	0.35 <sup>*</sup>	84.58	589.09
Santa Clara	13.35 <sup>*</sup>	22.24 <sup>*</sup>	0.25 <sup>*</sup>	87.76	6.67	1118.52 <sup>*</sup>	22.39 <sup>*</sup>	0.26 <sup>*</sup>	86.95	559.26
2016/2017										
Cordillera	39.68 <sup>*</sup>	13.54 <sup>*</sup>	0.13 <sup>*</sup>	101.12	19.84	4320.10 <sup>*</sup>	14.82 <sup>*</sup>	0.15 <sup>*</sup>	99.96	2160.05
Gaucha	18.94 <sup>*</sup>	16.04 <sup>*</sup>	0.17 <sup>*</sup>	92.42	9.47	2921.10 <sup>*</sup>	18.65 <sup>*</sup>	0.20 <sup>*</sup>	92.44	1460.55

<sup>1</sup> Abscissa of the inflection point.

<sup>2</sup> Ordinate of the inflection point.

\* Non-zero parameters according to the Student *t* test, *P* < 0.05.

**Table 4**

*P*-values for the F test to evaluate the identity of the logistic model between tomato genotypes, for the variables of number and weight of fruit (in grams) per plant, for 2015/2016 and 2016/2017.

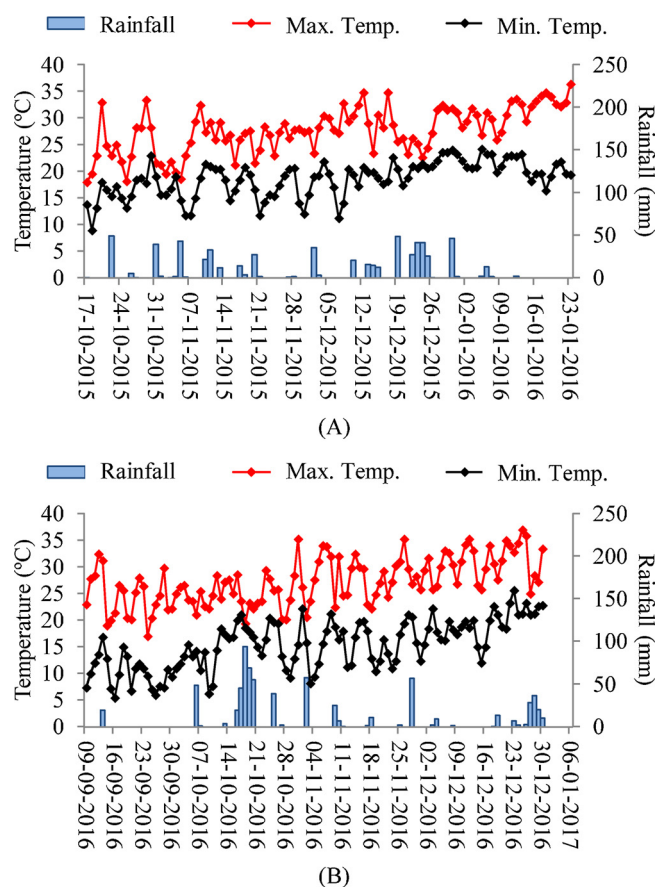
Hypotheses	Year	Number of fruits <i>p</i> -values	Weight of fruit
$\beta_1$			
$\beta_{11} = \beta_{12}$	2015/16	0.00	0.00
$\beta_{11} = \beta_{13}$	2015/16	0.00	0.00
$\beta_{12} = \beta_{13}$	2015/16	0.00	0.19
$\beta_{11} = \beta_{14}$	2016/17	0.00	0.00
$\beta_2$			
$\beta_{21} = \beta_{22}$	2015/16	0.72	0.88
$\beta_{21} = \beta_{23}$	2015/16	0.04	0.04
$\beta_{22} = \beta_{23}$	2015/16	0.22	0.07
$\beta_{21} = \beta_{24}$	2016/17	0.22	0.07
$\beta_3$			
$\beta_{31} = \beta_{32}$	2015/16	0.69	0.82
$\beta_{31} = \beta_{33}$	2015/16	0.03	0.03
$\beta_{32} = \beta_{33}$	2015/16	0.18	0.04
$\beta_{31} = \beta_{34}$	2016/17	0.08	0.03

\* $\beta_1$  = Parameters for the Cordillera genotype;  $\beta_2$  = parameters for the Ellen genotype;  $\beta_3$  = parameters for the Santa Clara genotype;  $\beta_4$  = parameters for the Gaucha genotype.

statistically higher in relation to the Santa Clara genotype (Table 3 and 4). Although the higher estimates values of  $\beta_2$  indicate otherwise, the Cordillera genotype produced fruits earlier (earlier occurrence of PI) (Fig. 1). For the 2015/2016 crop, the same is observed. Although the higher estimates values of  $\beta_2$  indicate otherwise (Table 3), the Gaucha genotype produced fruits earlier (Fig. 2). Therefore, the productive precocity of these genotypes is related to the higher estimate value of  $\beta_3$ .

The point of the cycle at which the rate of production is at its maximum is represented by the abscissa value for the inflection point ( $X_{IP}$ ). For the 2015/2016 crop,  $X_{IP}$  did not differ statistically between the Cordillera and Ellen genotypes for either variable. For these genotypes, the inflection point for both the weight and the number of fruits variables was reached at approximately 85 DAT. For the Santa Clara genotype, the inflection point was reached at a later stage for both variables, at around 87 DAT (Table 3 and Fig. 1). For the 2016/2017 crop, there was a statistical difference between  $X_{IP}$  values. The Gaucha genotype reached the inflection point for both variables at approximately 93 DAT, which was earlier than the Cordillera genotype, which reached the inflection point at around 100 DAT (Table 3 and Fig. 2).

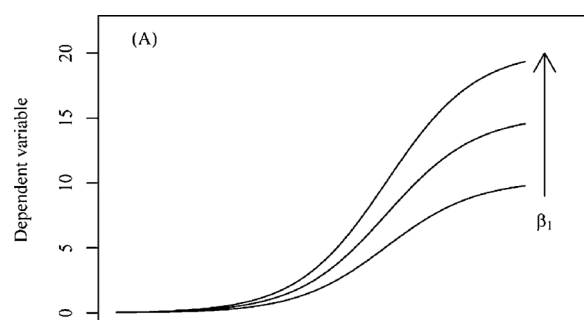
From simulated data, the relationship between lower estimates of  $\beta_2$  and higher estimation of  $\beta_3$  with the productive precocity of the genotypes was verified. The shift of the growth curve to the left (Fig. 4B), the higher slope (Fig. 4C) and the earlier occurrence of the IP (Fig. 5B



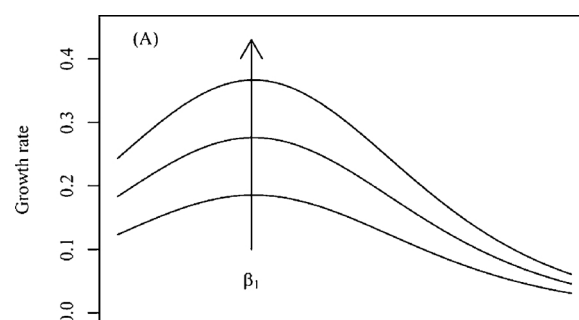
**Fig. 3.** Maximum temperature (Max. Temp.), minimum temperature (Min. Temp.) and daily rainfall for 2015/16 (A) and 2016/17 (B).

and C) were used in this interpretation. It is clear that the productive precocity, represented by the earliest occurrence of the PI (lower  $X_{IP}$ ), is related to the higher estimates of  $\beta_3$  in this study.

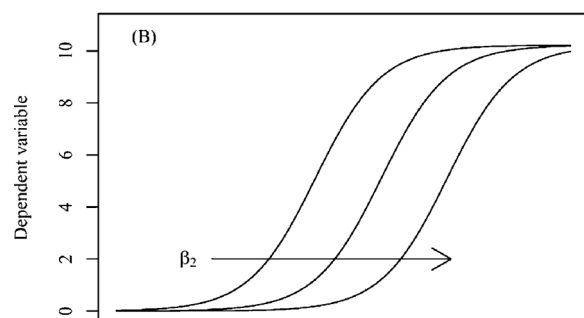
In addition to determining total production through the parameter  $\beta_1$ , the growth models increase the inferences that can be made through the interpretation of the other parameters (mainly  $\beta_3$  in this study) and the determination IP coordinates on the curve. Similarly, comparisons between nested models and estimated confidence intervals for the critical points allow treatments to be compared (Ritz and Streibig, 2008; Mischan et al., 2011; Pinho et al., 2014). This way it was possible to make inferences on the productive precocity of the tomato genotypes evaluated. In 2015/2016, the Cordillera and Ellen genotypes were more premature than the Santa Clara genotype. For the 2016/2017 crop, the



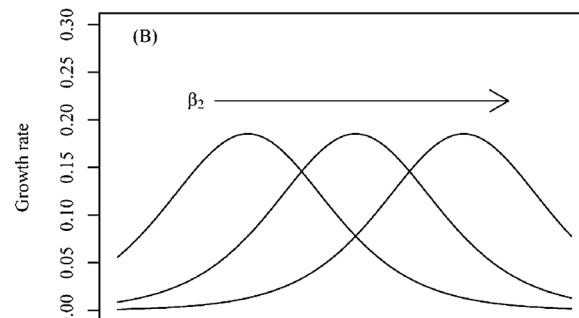
Independent variable



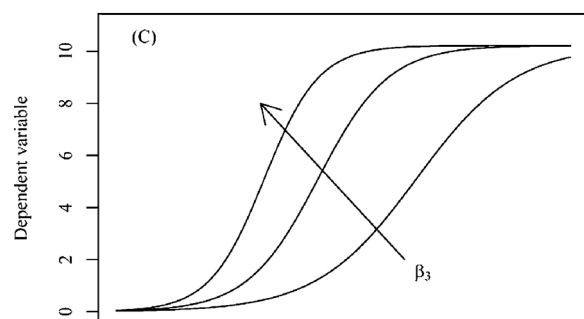
Independent variable



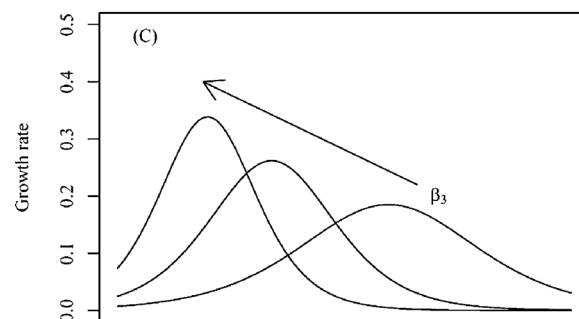
Independent variable



Independent variable



Independent variable



Independent variable

**Fig. 4.** Effect of increases in parameters  $\beta_1$  (A),  $\beta_2$  (B) and  $\beta_3$  (C) on the logistic model:  $Y_i = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 t_i)}} + \varepsilon_i$ . Arrows indicate changes in the curves as the parameter values increase, all other constants remaining the same.

Gaucha genotype tended to be more premature than the Cordillera genotype (Tables 3 and 4 and Figs. 2 and 3).

### 3.4. Use of growth models as a statistical analysis tool

For the statistical analysis of the experiments, the use of the logistic model determined total production for the genotypes and enabled comparisons to be made on other aspects of the productive behavior of the tomato plants. However, growth models can be used in other ways,

**Fig. 5.** Effect of increases in parameters  $\beta_1$  (A),  $\beta_2$  (B) and  $\beta_3$  (C) on the growth rate in the logistic model:  $Y_i = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 t_i)}} + \varepsilon_i$ . Arrows indicate changes in the curves as the parameter values increase, all other constants remaining the same.

and not just as a statistical analysis tool. For example, they can be used to get a better understanding of the productive behavior of vegetables over harvest periods and describe this through growth curves (and their derivatives), can be used to help researchers or agricultural extension agents to recommend planting periods for genotypes to avoid overloading workforces during determined points of the harvest period, or to plan production in terms of commercialization (Figs. 2 and 3). Wider studies, including tests into genotype and transplant period, can contribute towards technical recommendations for crops.

Growth models have proven to be efficient statistical analysis tools. They have been used to make inferences regarding genotype productive behavior based on yield data that can be easily collected in the field and reported according to the number of days after harvesting. Furthermore, model parameters and critical points (especially  $\beta_3$  and  $X_{IP}$ ) can be used to identify the genotypes with the highest yield, predict the time of highest yield and therefore the time at which the yield is greatest. In other words, productive behavior can be quantified by the model's parameter values and its inflection point, which can be used as explanatory variables to compare genotypes. This information is important but is largely ignored due to measurement problems. We have shown that using the correct statistical approach, these problems can be overcome.

Note that the approach proposed is not related to deterministic models, like TOMGRO (Jones et al., 1991). What we are proposing is to use growth models to model the cumulative sigmoidal yield behavior as a function of time after transplanting the seedlings. This approach is a workaround for the problems of ANOVA assumption violation and increases the inferences that can be made regarding tomato productive behavior, which can be compared based on the model's parameter values and critical points. Thus, we are proposing that ANOVA is replaced by comparison of models. Treatments, genotypes, and environments can be used as factors, and the significance of their effect tested by using nested models based on the F test.

To obtain the nonlinear parameters estimative by OLS, the models must be approximately linear (Seber and Wild, 2003; Bates and Watts, 2007). Thus, the low nonlinearity measures indicate that the parameter estimate is close to begin unbiased. As the parameter estimates have a biological interpretation and determine the IP and its confidence interval, the selection of the models must be performed by the low nonlinearity measures (that indicate close to begin unbiased estimates). In this way, the model's selection must be performed by these measures, and not only by the numerical statistics that measure a goodness of fit (e.g.,  $R^2$ ) or selection criteria measures (e.g., AIC, BIC). Using only the goodness of fit measures, the models selected can have wrong.

In conclusion, the models with a higher goodness of fit can be having parameters estimates not consistent with the reality observed. Thus, the nonlinearity assess is fundamental. Due to this, the selected model in this study was performed based on the lower nonlinearity measures (Table 2 and 3). The  $R^2$  was estimated to verify that the selected model had a high goodness of fit.

### 3.5. Defining productive precocity

This work proposed a definition of precocity based on the productive behavior of tomato genotypes. We define the productive precocity as the moment of PI occurrence because the farmer has economies of scale at this moment. In addition, we provide statistical bases (for the model's selection and parameters interpretation) for this characteristic (precocity) to be determined. Thus, the productive precocity of tomato genotypes can be determined with statistical rigor based on their productive behavior, and not subjectively.

Therefore, growth models can be used to compare genotypes or treatments of multiple harvest vegetables. Given the biological interpretation of the parameters, both final production and other aspects relating to production can be compared, discussed and analyzed, increasing the inferences that can be made regarding the productive behavior of other multiple harvest crops. Know the economies of scale moment are important for the farmer to plan the vegetable crop harvests, both from the practical point of view (sizing of the workforce) and from the financial point of view (planning harvests for periods with better prices).

The use of the nonlinear regression to determine a precocity production can advantages in relation to the precocity index (such as proposed by Khanizadeh and Fanous, 1992): i) it is possible to visualize graphically the increase of production by the first order derivative of the

curves (Fig. 1 and 2); ii) it is possible to identify the exact moment at the maximum rate production occur, and this moment indicates the productive precocity; iii) the parameters estimate  $\beta_2$  and  $\beta_3$  can be used with an index to determine the precocity of production in multiple harvests crops.

## 4. Conclusion

The growth models made possible describe the behavior of the tomato plant production. The selection of the logistic model was performed by the lower nonlinearity measures. The  $\beta_1$  estimates represent the total production of the genotypes. By the  $\beta_3$  estimate and of the inflection point, it was possible to determine the genotypes production precocity. In this study, the Cordillera and Ellen genotypes were more premature in 2015/2016 crop, and the Gaucho genotype was more premature in the 2016/2017 crop.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.scienta.2018.10.044>.

## 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>.
- Bem, C.M., Filho, A.C., Facco, G., Schabarum, D.E., Silveira, D.L., Simões, F.M., Uliana, D.B., 2017v. Growth models for morphological traits of sunn hemp. *Semina: Ciênc. Agrár.* 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., 2018. Gompertz and logistic models to the productive traits of Sunn hemp. *J. Agric. Sci.* 10, 225–238. <https://doi.org/10.5539/jas.v10n1p225>.
- Bates, D.M., Watts, D.G., 2007. *Nonlinear Regression Analysis and Its Applications*. John Wiley and Sons Inc., New York, pp. 392.
- Fernandes, T.J., Pereira, A.P., Muniz, J.A., 2017. Double sigmoidal models describing the growth of coffee berries. *Cienc. Rural* 47, e20160646. <https://doi.org/10.1590/0103-8478cr20160646>.
- Fernandez, G.C.J., 2007. Design and analysis of commonly used comparative horticultural experiments. *HortScience* 42, 1052–1069.
- Godoy, C., Monterubbianesi, G., Tognetti, J., 2008. Analysis of highbush blueberry (*Vaccinium corymbosum* L.) fruit growth with exponential mixed models. *Sci. Hortic.* 193, 286–293. <https://doi.org/10.1016/j.scienta.2007.10.018>.
- IBGE – Instituto Brasileiro de Geografia e Estatística, 2017. *Levantamento sistemático da produção agrícola*, março de 2017. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, pp. 83.
- Jones, J.W., Dayan, E., Allen, L.H., Van Keulen, H., Challa, H., 1991. A dynamic tomato growth and yield model (TOMGRO). *ASAE* 34, 663–672. <https://doi.org/10.13031/2013.31715>.
- Khanizadeh, S., Fanous, M.A., 1992. Mathematical indices for comparing small fruit crops for harvest time and trait similarity. *HortScience* 27, 346–348.
- Lúcio, A.D., 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., Rodrigues, M., Bevilacqua, L.M., Voss, H.M.G., Copetti, D., Faé, M., 2016a. Modelos não-lineares para a estimativa da produção de tomate do tipo cereja. *Cienc. Rural* 46, 233–241. <https://doi.org/10.1590/0103-8478cr20150067>.
- Lúcio, A.D., Nunes, L.F., Rego, F., 2016b. Nonlinear regression and plot size to estimate green beans production. *Hortic. Bras.* 34, 507–513. <https://doi.org/10.1590/s0102-053620160409>.
- 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>.
- Mischan, M.M., Pinho, S.Z., 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>.
- Muniz, J.A., Nascimento, M.S., Fernandes, T.J., 2017v. Nonlinear models for description of cacao fruit growth with assumption violations. *Rev. Caatinga* 30, 250–257. <https://doi.org/10.1590/1983-21252017v30n128rc>.
- Paine, C.E.T., Marthews, T.R., Vogt, D.R., Purves, D., Rees, M., Hector, A., Turnbull, L.A., 2012. How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods Ecol. Evol.* 3, 245–256. <https://doi.org/10.1111/j.2041-210X.2011.00155.x>.
- Pinho, S.Z., Carvalho, L.R., Mischan, M.M., Passos, J.R.S., 2014. Critical points on growth curves in autoregressive and mixed models. *Sci. Agric.* 71, 30–37. <https://doi.org/10.1590/S0103-90162014000100004>.
- Puiatti, G.A., Cecon, P.R., Nascimento, M., Puiatti, M., Finger, F.L., Silva, A.R., Nascimento, A.C.C., 2013. *Análise de agrupamento em seleção de modelos de*



- regressão não lineares para descrever o acúmulo de matéria seca em plantas de alho. *Rev. Bras. Biom.* 31, 337–351.
- R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Austria.
- Ribeiro, T.D., Savian, T.V., Fernandes, T.J., Muniz, J.A., 2018. The use of the nonlinear models in the growth of pears of ‘Shinseiki’ cultivar. *Cienc. Rural* 47, e20161097. <https://doi.org/10.1590/0103-8478cr20161097>.
- Ritz, C., Streibig, J.C., 2008. *Nonlinear Regression with R*. Springer, New York, pp. 142.
- ROLAS - Rede Oficial de Análise de Solo e de Tecido Vegetal, Manual de adubação e calagem para os estados do Rio Grande do Sul e Santa Catarina, 2004. Sociedade Brasileira De Ciência Do Solo, Porto Alegre. pp. 400.
- Seber, G.A.F., Wild, C.J., 2003. *Nonlinear Regression*. John Wiley and Sons Inc, New Jersey, pp. 768.
- Sousa, I.F., Neto, J.E.K., Muniz, J.A., Guimarães, R.M., Savian, T.V., Muniz, F.R., 2014. Fitting nonlinear autoregressive models to describe coffee seed germination. *Cienc. Rural* 44, 2016–2021. <https://doi.org/10.1590/0103-8478cr20131341>.
- Wubs, A.M., Ma, Y.T., Heuvelink, E.P., 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.
- Zeileis, A., Hothorn, T., 2002. Diagnostic checking in regression relationships. *R News* 2, 7–10.