

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIENCIAS RURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

Bruno Giacomini Sari

**PARÂMETROS BIOLÓGICOS DA PRODUÇÃO DE TOMATEIRO VIA
MODELO LOGÍSTICO**

Santa Maria, RS
2018

Bruno Giacomini Sari

**PARÂMETROS BIOLÓGICOS DA PRODUÇÃO DE TOMATEIRO VIA MODELO
LOGÍSTICO**

Tese apresentada ao Programa de Pós Graduação em Agronomia, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do título de **Doutor em Agronomia**.

Orientador: Prof. Dr. Alessandro Dal'Col Lúcio

Santa Maria, RS
2018

Sari, Bruno Giacomini
PARÂMETROS BIOLÓGICOS DA PRODUÇÃO DE TOMATEIRO VIA
MODELO LOGÍSTICO / Bruno Giacomini Sari.- 2018.
125 p.; 30 cm

Orientador: Alessandro Dal'Col Lúcio
Tese (doutorado) - Universidade Federal de Santa
Maria, Centro de Ciências Rurais, Programa de Pós
Graduação em Agronomia, RS, 2018

1. Modelos de crescimento 2. Regressão não linear 3.
Precocidade produtiva 4. Concentração de colheitas 5.
Solanum lycopersicum I. Lúcio, Alessandro Dal'Col II.
Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

© 2018

Todos os direitos autorais reservados a Bruno Giacomini Sari. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte.

Endereço: Rua Sete de Setembro, n. 764, Centro, Faxinal do Soturno, RS. CEP: 97220-000

E-mail:brunosari@hotmail.com

ERRATA

Página	Linha	Onde se lê	Leia-se
5	1	Bruno GiacominiSari	Bruno Giacomini Sari
5	13	José AntônioGonzalez da Silva	José Antônio Gonzalez da Silva

Bruno Giacominisari

PARÂMETROS BIOLÓGICOS DA PRODUÇÃO DE TOMATEIRO VIA MODELO LOGÍSTICO

Tese apresentada ao Programa de Pós Graduação em Agronomia, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do título de **Doutor em Agronomia.**

Aprovado em 09 de março de 2018:



Alessandro Dal'Col Lúcio, Dr. (UFSM)
(Presidente/Orientador)



Alberto Cargnelutti Filho, Dr. (UFSM)



José Antônio Gonçalves da Silva, Dr. (Unijuí)



Sidinei José Lopes, Dr. (UFSM)



Taciana Villela Sayfan, Dr^a. (USP/ESALQ) - Videoconferência

Santa Maria, RS
2018

DEDICATÓRIA

*A minha esposa Ivete Streck Sari e a nossa filha Julia Streck Sari, sem o amor e a compreensão de vocês nada seria possível.
Dedico também a minha mãe Maria Inês Giacomini e a meu pai Darci Carlos Sari.*

AGRADECIMENTOS

À DEUS por tudo.

À UNIVERSIDADE FEDERAL DE SANTA MARIA pelo acolhimento e ao PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA pela oportunidade de cursar o mestrado e o doutorado.

À COORDENAÇÃO DE APERFEIÇOAMENTO PESSOAL DE NÍVEL SUPERIOR (CAPES) pela concessão da bolsa de estudo tanto no mestrado quanto no doutorado.

Ao professor ALESSANDRO DAL'COL LÚCIO, pela oportunidade de cursar o mestrado e doutorado junto ao seu grupo de pesquisa, pela sua dedicação, pelos seus ensinamentos e por sua amizade. Obrigado por oportunizar o meu crescimento pessoal e profissional. Foi um prazer ter todos esses anos de convivência e aprendizado!

À professora TACIANA VILLELA SAVIAN, por suas contribuições e observações na prova de qualificação de doutorado. A sua contribuição permitiu a realização deste trabalho.

À minha esposa IVETE STRECK SARI e a nossa a nossa filha JULIA STRECK SARI. A minha esposa pelo apoio incondicional, pois sem ele esse trabalho não se realizaria. A minha filha por me alegrar e me “encher de vida” nos momentos difíceis. As dificuldades por nós enfrentadas durante esse doutoramento provou o quanto a nossa família é unida e o nosso amor é incondicional!.

À toda a minha FAMÍLIA, em especial a minha mãe MARIA INÊS GIACOMINI, ao meu pai DARCI CARLOS SARI e aos meus irmãos THIAGO GIACOMINI SARI, MARCELO GIACOMINI SARI e LAURA GIACOMINI SARI. Como é bom ter vocês por perto.

À minha avó, AURELIA SPONCHIADO SARI, por compreender a minha ausência. A dificuldade da senhora no passado que permitiu que eu estudasse e pudesse concluir esse doutorado.

Aos colegas, amigos e FUNCIONÁRIOS DO DEPARTAMENTO DE FITOTECNIA pelo apoio na condução dos experimentos e pelos anos de convivência e amizade.

Aos alunos de graduação e estagiários do grupo de pesquisa EXPERIMENTAÇÃO, ANDRÉ LUÍS TISCHLER, LUCAS DREBES, ANDRIELLE MAGRINI e LETICIA GALL DREIFKE pelo apoio na condução dos experimentos.

Aos colegas do grupo de pesquisa EXPERIMENTAÇÃO, em especial aos da “salinha da pós” CINTHYA SOUZA SANTANA, DIONATAN KETZER KRYSCZUN, MARIA INÊS DIEL e TIAGO OLIVOTO. Obrigado pela amizade e pela convivência, frequentar aquela sala com vocês sempre foi um momento de muita alegria.

Não poderia de fazer um agradecimento especial ao colega TIAGO OLIVOTO, pelo auxílio na tradução dos artigos para o inglês e na configuração das figuras desta tese. Valeu Tiagão!

Ao meu grande amigo e compadre JOELTON DOS SANTOS RODRIGUES, pela amizade que construímos ao longo da graduação e temos até hoje. Grande abraço Joeltão!

À todos, MUITO OBRIGADO!

RESUMO

DESCRIÇÃO DA PRODUÇÃO DE TOMATEIRO POR MODELOS NÃO LINEARES

AUTOR: Bruno Giacomini Sari
ORIENTADOR: Alessandro Dal'Col Lúcio

As olerícolas de múltiplas colheitas caracterizam-se por apresentarem elevada variabilidade entre plantas e colheitas. Esse aspecto limita o uso da ANOVA, pois os pressupostos de normalidade, homogeneidade e independência dos resíduos são violados. Porém, existem inúmeras abordagens estatísticas que podem ser utilizadas para analisar experimentos com olerícolas de múltiplas colheitas, entre estes, os modelos não lineares. Portanto, o presente trabalho tem como objetivo descrever o comportamento produtivo do tomate ao longo do tempo (colheitas) através de modelos não lineares e indica-los como ferramenta de análise estatística para ser usado em experimentos com tomate. Dois experimentos (2015/2016 e 2016/2017) foram conduzidos a campo no departamento de Fitotecnia da Universidade Federal de Santa Maria (UFSM) com diferentes genótipos de tomate: Cordillera, Ellen e Santa Clara no primeiro ano; e Cordillera e Gaucho no segundo. Os frutos foram colhidos semanalmente, sendo eles contados e pesados. As variáveis número e massa de frutos foram acumuladas consecutivamente (a cada colheita), e a estes dados foram ajustados os modelos Brody, Gompertz, logístico e von Bertalanffy. O melhor modelo foi selecionado com base no valor do coeficiente de determinação (R^2) e na não linearidade paramétrica. A igualdade entre os parâmetros do modelo selecionado, estimados em cada genótipo, foi verificada utilizando o teste F. Por fim, foram obtidos os pontos críticos do modelo: ponto de aceleração máximo, ponto de inflexão, ponto de desaceleração máximo e ponto de desaceleração assintótico. As variáveis apresentaram comportamento sigmoidal, o que possibilitou o ajuste dos modelos de crescimento. Entre os modelos testados, foi selecionado aquele com elevada capacidade de predição e não linearidade baixa, indicando que as estimativas são aproximadamente não viesadas. Com base nas estimativas dos parâmetros e nos pontos críticos do modelo selecionado, foram realizadas inferências quanto a produção total, precocidade produtiva e concentração da produção. O modelo logístico foi o selecionado, independente do ano ou do genótipo, pois apresentou baixa não linearidade paramétrica, e com base nas estimativas dos seus parâmetros e pontos críticos foi possível realizar inferências quanto ao comportamento produtivo dos genótipos. No primeiro ano, o genótipo Cordillera foi o mais produtivo e o mais precoce, atingindo o pico de produção aproximadamente aos 85 dias após o transplante das mudas (DAT), com uma produção concentrada entre os 82 e 89 DAT, aproximadamente. O genótipo Ellen foi o menos produtivo, porém o seu comportamento foi semelhante: a produção se concentrou entre os 82 e 89 DAT, com pico aos 85 DAT. O genótipo Santa Clara obteve uma produção intermediária e pouco precoce, uma vez que atingiu o pico de produção apenas aos 90 DAT, e concentrou sua produção entre os 92 e 102 DAT. No segundo ano, o genótipo Gaucho apresentou uma produção menor, porém mais concentrada que o genótipo Cordillera. No genótipo Gaucho o pico da produção foi observado aproximadamente aos 90 DAT (concentrando-se entre 85 e 100 DAT, aproximadamente), enquanto que o genótipo Cordillera o pico ocorreu por volta dos 100 dias (concentrando-se entre 92 e 110 DAT, aproximadamente). Os modelos de crescimento mostraram-se uma alternativa de análise estatística de experimentos com tomate. Além disso, a partir dos seus parâmetros e pontos críticos, é possível realizar inferências quanto a produção, precocidade produtiva e a concentração da produção de frutos de tomate. Apesar desse trabalho se concentrar na cultura do tomate, os modelos são uma alternativa de análise para qualquer cultura olerícola.

Palavras-chave: *Solanum lycopersicum*. Modelos de crescimento. Precocidade produtiva. Olerícolas de múltiplas colheitas.

ABSTRACT

DESCRIBING TOMATO PLANT PRODUCTION USING NONLINEAR MODELS

AUTHOR: Bruno Giacomini Sari
ADVISOR: Alessandro Dal'Col Lúcio

The multiple harvest vegetable crops are characterized by high variability between plants and crops. This characteristic limits the use of ANOVA, since the assumptions of normality, homogeneity and independence of residues are violated. However, there are a number of statistical approaches that can be used to analyze experiments with multiple harvested crops, between them nonlinear models. Therefore, this work aims to describe the productive behavior of the tomato over time (harvests) through non-linear models and indicates them as a statistical analysis tool to be used in tomato experiments. Two experiments (2015/2016 and 2016/2017) were conducted in the field in the Crop Science Department of the Federal University of Santa Maria (UFSM) with different tomato genotypes: Cordillera, Ellen and Santa Clara in the first year; and Cordillera and Gaucho in the second year. The fruits were harvested weekly, counted and weighed. The number and mass of fruits were consecutively accumulated (at each harvest), and the Brody, Gompertz, Logistic and von Bertalanffy models were fitted to these data. The best model was selected based on the value of the coefficient of determination (R^2) and on the parametric nonlinearity. Finally, the critical points of the selected model were obtained: maximum acceleration point, inflection point, maximum deceleration point, and asymptotic deceleration point. The variables showed sigmoidal behavior, which allowed the adjustment of the growth models. Among the models tested, it was selected the one with high prediction capacity and low nonlinearity, indicating that estimates are approximately unbiased. Based on the estimates of the parameters and the critical points of the selected model, inferences were made regarding total production, productive precocity and concentration of production. The logistic model was selected, independent of the year or genotype, because it presented low parametric nonlinearity, and based on the estimations of its parameters and critical points it was possible to make inferences regarding the productive behavior of the genotypes. In the first year, the Cordillera genotype was the most productive and the most precocious, reaching the peak of production approximately 85 days after the transplanting of the seedlings (DAT), with a concentrated production between approximately 82 and 89 DAT. The Ellen genotype was the least productive, but its behavior was similar: production was concentrated between 82 and 89 DAT, with a peak at 85 DAT. The Santa Clara genotype obtained an intermediate and early production, since it reached the peak of production only at 90 DAT, and concentrated its production between 92 and 102 DAT. In the second year, the genotype Gaucho presented a smaller production, but more concentrated than the Cordillera genotype. In the Gaucho genotype peak production was observed at approximately 90 DAT (concentrating between approximately 85 and 100 DAT), while the Cordillera peak genotype occurred around 100 days (concentrating between 92 and 110 DAT, approximately). The growth models proved to be an excellent alternative for the statistical analysis of experiments with multiple harvested olive groves. In addition, from its parameters and critical points, it is possible to make inferences about production, precocity and concentration of production. Although this work focuses on tomato culture, the models can be an alternative analysis for any olive cultivation.

Keywords: *Solanum lycopersicum*. Growth models. Productive precocity. Multiple harvest vegetable crops.

LISTA DE GRÁFICOS

REVISÃO BIBLIOGRÁFICA

- Figura 1 - Diferentes tipos de zeros observados em culturas olerícolas de múltiplas colheitas. 28
- Figura 2 - Porcentagens de zero e coeficiente de variação (CV%) entre plantas por colheita para a variável número de frutos por planta nos genótipos de tomate Gaúcho (A), Cordillera (B) e Janaína (C). Fonte: dados pessoais. 29
- Figura 3 - Ajuste do modelo logístico (A), taxa de produção de frutos (B) e taxa de aceleração (C) dos genótipos Cordillera e Gaucho Fonte: dados pessoais. 31
- Figura 4 - Contornos da soma de quadrados dos resíduos e estimativas dos parâmetros obtidos por 5000 reamostragens bootstrap do modelo Gompertz e logístico ajustado a dados de produção de berinjela. Fonte: dados pessoais. 36
- Figura 5 - QQ-Plot dos resíduos do modelo logístico ajustado a dados de produção (massa de frutos) do genótipo de tomate Cordillera, colhido a cada três e sete dias. Fonte: dados pessoais. 39
- Figura 6 - Gráfico dos resíduos (simulados) vs colheitas indicando heterocedasticidade e homocedasticidade dos resíduos. Fonte: dados pessoais. 41
- Figura 7 - Gráfico quantil-quantil, histogramas de frequência e box-plot em condições de normalidade (A,B), assimetria a esquerda (C,D), assimetria a direita (E,F), distribuição leptocúrtica (G,H) e distribuição platicúrtica (I,J), utilizando dados simulados. 42
- Figura 8 - Gráficos de autocorrelação (ACF) e autocorrelação parcial (PAF) obtidos a partir de resíduos correlacionados (A,B) e independentes (C,D) do modelo logístico ajustado a dados de produção de frutos de morangueiro. Fonte: dados pessoais. ... 43
- Figura 9 - Alteração (indicada pela seta) no modelo logístico (A), na derivada de primeira (B) e segunda ordem (C) com o incremento dos parâmetros β_1 55
- Figura 10 - Alteração (indicada pela seta) no modelo logístico (A), na derivada de primeira (B) e segunda ordem (C) com o incremento dos parâmetros β_2 56
- Figura 11 - Alteração (indicada pela seta) no modelo logístico (A), na derivada de primeira (B) e segunda ordem (C) com o incremento dos parâmetros β_3 57

ARTIGO I

- Figura 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. 70
- Figura 2 - Observed values (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 2016/2017. 71
- Figura 3 - Maximum temperature (Max. Temp.), minimum temperature (Min. Temp.) and daily rainfall for 2015/16 (A) and 2016/17 (B). 75
- Figura 4 - Effect of increases in parameters β_1 (A), β_2 (B) and β_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. 77

Figura 5 - Effect of increases in parameters β_1 (A), β_2 (B) and β_3 (C) on the growth rate in the logistic model: $Y_i = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 t)}} + \varepsilon_i$. Arrows indicate changes in the curves as the parameter values increase, all other constants remaining the same. 78

ARTIGO II

Figura 1 - Parameters of the logistic model for the number (A) and mass (B) of fruits per plant in three tomato genotypes grown in 2015/2016	97
Figura 2 - Parameters of the logistic model for the number (A) and mass (B) of fruits per plant in three tomato genotypes grown in 2016/2017.	98
Figura 3 - Logistic model (A), fruit production rate (B) and growth acceleration (C) fitted to the number of fruits of three tomato genotypes in 2015/2016 GS.	99
Figura 4 - 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.....	100
Figura 5 - 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.	101
Figura 6 - 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.....	102
Figura 7 - Maximum air temperature, minimum air temperature and daily precipitation in the 2015/16 (A) e 2016/17 (B) GSs.....	104

LISTA DE TABELAS

REVISÃO BIBLIOGRÁFICA

Tabela 1 - Testes estatísticos recomendados para testar os pressupostos e suas respectivas funções no software R.....	40
Tabela 2 - Função, velocidade e aceleração dos modelos logístico, Brody, Gompertz e von Bertalanffy.	48
Tabela 3 - Coordenadas [x;y] dos pontos de aceleração máxima (PAM), inflexão (PI), desaceleração máxima (PDM) e desaceleração assintótica (PDA) obtidos matematicamente a partir dos modelos logístico, Brody, Gompertz e von Bertalanffy (MISCHAN; PINHO, 2014).....	50

ARTIGO I

Tabela 1 - P-values for the Shapiro-Wilk (SW), Breusch-Pagan (BP) and Durbin-Watson (DW) 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.	69
Tabela 2 - P-values for the Shapiro-Wilk (SW), Breusch-Pagan (BP) and Durbin-Watson (DW) 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 2016/2017.	69
Tabela 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.	73
Tabela 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.	74

ARTIGO II

Tabela 1 - P-values for Shapiro-Wilk (SW), Breusch-Pagan (BP) and Durbin-Watson (DW) tests, determination coefficient (R^2), and nonparametric linearity (c^0) 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.....	95
Tabela 2 - P-values for Shapiro-Wilk (SW), Breusch-Pagan (BP) and Durbin-Watson (DW) tests, determination coefficient (R^2), and nonparametric linearity (c^0) 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.....	96

DISCUSSÃO GERAL

Tabela 1 - Número médio de frutos por planta e concentração de produção de frutos estimados pelo modelo logístico em cada genótipo de tomate nos dois anos de condução dos experimentos.	114
---	-----

Discussion	103
<i>Production and climate variables</i>	103
<i>Models' fit and assumptions</i>	105
<i>Biological interpretation of critical points of the models</i>	106
<i>Precocity and concentration of production</i>	108
Conclusions	109
Acknowledgement	110
References	110
5 DISCUSSÃO GERAL	113
6 CONCLUSÕES	117
REFERÊNCIAS	119
ANEXO A – CARACTERÍSTICAS DOS GENÓTIPOS DE TOMATE	125

1 INTRODUÇÃO

Devido a importância econômica do tomate, estudos que geram informações ao agricultor são extremamente importantes. Porém, a qualidade dessas informações dependerá da precisão estatística dos experimentos conduzidos com a cultura. No caso das ciências agrárias, grande parte dos dados gerados em experimentos é analisada utilizando a análise da variância (ANOVA). Uma das características deste tipo de análise é a necessidade da aderência dos resíduos a normalidade, que eles sejam independentes e homocedásticos (CASELLA, 2008; RENCHER; SCHAALJE, 2008).

Em experimentos com olerícolas, esses pressupostos geralmente são violados pois a distribuição das variáveis produtivas (massa e número de frutos) não possui distribuição normal, há elevada variabilidade entre plantas dentro de uma mesma colheita e o banco de dados é altamente inflacionado por zeros. Isso ocorre devido a características da cultura, principalmente a maturação desuniforme e o ponto subjetivo das colheitas (CARGNELUTTI FILHO et al., 2004; LÚCIO; SARI, 2017), podendo causar violações sistemáticas dos pressupostos dos modelos da ANOVA (LÚCIO et al., 2012b).

Apesar das variáveis produtivas do tomate não terem distribuição normal (geralmente são distribuições inflacionadas de zero nas colheitas individuais), quando as colheitas e/ou plantas são acumuladas observa-se uma tendência destas variáveis se aproximarem a esta distribuição (COUTO et al., 2009; LÚCIO et al., 2010). Isso leva a uma maior possibilidade de cumprimento dos pressupostos dos modelos da ANOVA, viabilizando a sua utilização. Por isso, o pesquisador tende a utilizar parcelas grandes e analisar apenas a produção total (LÚCIO et al., 2012; LÚCIO; SARI, 2017). Porém, quando o pesquisador lança mão desta técnica há uma perda de grande quantidade de informações a respeito da produção da cultura ao longo do ciclo produtivo da mesma.

A utilização da ANOVA por parte dos pesquisadores ocorre muitas vezes pela impossibilidade (ou desconhecimento) de utilizar novas abordagens estatísticas. Porém, com o aumento da capacidade computacional e os softwares livres, a utilização de modelos antes pouco utilizados passou a ser facilitado. No caso de olerícolas com medidas repetidas no tempo, as múltiplas colheitas, os modelos mistos ou modelos lineares generalizados são alternativas mais eficientes do que a ANOVA (FERNANDEZ, 2007).

Ressalta-se que a ANOVA é uma ferramenta de análise estatística importante, e sua utilização dependerá dos objetivos do estudo. Nem sempre as variáveis em estudo são inflacionadas com zero ou possuem superdispersão. Além disso, nem sempre conhecer o

comportamento produtivo é o objetivo do pesquisador. Nestes casos a utilização da ANOVA é perfeitamente cabível. Porém em estudos que visem conhecer o comportamento produtivo das olerícolas ao longo do tempo (presença de medidas repetidas), outros métodos surgem como alternativa à ANOVA. Como esse trabalho se concentra na busca de alternativas de análises do comportamento produtivo ao longo do tempo, foram encontradas metodologias para solucionar as fragilidades observadas na ANOVA neste tipo de análise.

Uma abordagem extremamente interessante para analisar experimentos com olerícolas é a utilização de modelos de regressão não lineares. Como o comportamento das variáveis acumuladas de massa e número de frutos é sigmoide, a utilização de modelos de crescimento é recomendada para a sua modelagem (SEBER; WILD, 2003). Esses modelos empíricos possuem parâmetros com interpretação biológica que podem ser utilizados para descrever a produção de olerícolas ao longo do tempo.

A utilização dos modelos deve levar em considerações aspectos que permitam utilizar os parâmetros dos modelos como variáveis explicativas do comportamento produtivo. O mais importante destes aspectos é o grau de não linearidade do modelo, que deve ser baixo para que os estimadores sejam aproximadamente não viesados (BATES; WATTS, 1988; SEBER; WILD, 2003). Nesse sentido, a simples seleção de modelos pela qualidade de ajuste (maior R^2 , menor AIC, entre outros) pode levar a resultados desastrosos, quando a não linearidade não é avaliada.

O uso de regressões não lineares ainda é incipiente em culturas olerícolas e limita-se a apenas ajustar os modelos aos dados acumulados de massa e frutos (LÚCIO et al., 2015; LÚCIO et al., 2016a; LÚCIO et al., 2016b). Porém a sua utilização pode ser mais abrangente, podendo aumentar as inferências que podem ser feitas quanto ao comportamento produtivo das olerícolas ao longo do tempo, utilizando para isso os parâmetros e os pontos críticos dos modelos. Percebe-se que as limitações da ANOVA exigem que novas técnicas sejam adotadas para analisar experimentos com olerícolas de múltiplas colheitas e que com a evolução computacional vários métodos estatísticos surgem como alternativa. Os modelos não lineares é uma delas.

A adoção de modelos não lineares como alternativa na análise estatística de dados provenientes de experimentos com essas características proporcionam recomendações mais detalhadas e com alto grau de confiabilidade. Percebe-se que tanto as estimativas dos parâmetros dos modelos quanto o momento de ocorrência dos pontos críticos determinam informações sobre a produção total, precocidade produtiva e o período em que a produção é concentrada. Essa abordagem, além de resolver problemas comuns em análise estatística de

olerícolas utilizando a ANOVA (heterocedasticidade e dependência dos resíduos), aumenta a inferência que se pode fazer quanto ao comportamento da produção das olerícolas ao longo do tempo. O benefício do uso dos modelos de crescimento não é apenas estatístico, mas também prático do ponto de vista fitotécnico. Para que os modelos atinjam esses resultados, importantes aspectos estatísticos devem ser cumpridos no momento de ajustar os modelos, principalmente em relação a estimativas aproximadamente não viesados¹ (RATKOWSKY, 1993).

Uma revisão bibliográfica dará o suporte para que os modelos não lineares sejam utilizados corretamente e, conseqüentemente, que os resultados sejam precisos. Posteriormente serão apresentados dois artigos, onde modelos serão ajustados a dados acumulados de massa e frutos de tomate e, a partir das estimativas dos seus parâmetros e dos pontos críticos, serão realizadas inferências quanto a produção, a precocidade produtiva e a concentração da produção em diferentes genótipos.

1.1 PROBLEMA

Como contornar os problemas de elevada variabilidade entre plantas e colheitas, que geram resultados imprecisos quando a ANOVA é utilizada, e descrever o comportamento produtivo da cultura do tomate ao longo do tempo?

1.2 OBJETIVOS

1.2.1 Objetivo geral

Descrever o comportamento produtivo do tomate ao longo do ciclo produtivo (múltiplas colheitas) via uso de modelos não lineares e indicá-los como ferramenta de análise estatística para ser adotado em experimentos com tomate.

1.2.2 Objetivos específicos

¹ Utilizou-se ao longo do texto a expressão “estimativa aproximadamente não viesada” para caracterizar os parâmetros estimados de modelos não lineares com uma boa aproximação linear. Essa é uma tradução livre de “the estimates were close to being unbiased” utilizada por Ratkowsky (1993) no seu artigo “Principles of nonlinear regression modeling”.

- Verificar se o comportamento produtivo das variáveis massa e número de frutos acumulados permite utilizar modelos de crescimento para descrevê-las;
- Selecionar entre um conjunto de modelos de crescimento qual o que melhor se ajusta as variáveis acumuladas de massa e número frutos de tomate;
- Utilizar as estimativas dos parâmetros do modelo selecionado, que possuem interpretação biológica, como variáveis que descrevem o comportamento produtivo do tomate;
- Utilizar as estimativas dos pontos críticos do modelo selecionado para descrever o comportamento produtivo do tomate;
- Fazer inferências quanto a produção, a precocidade produtiva e a concentração de produção de genótipos de tomate.

2 REVISÃO BIBLIOGRÁFICA

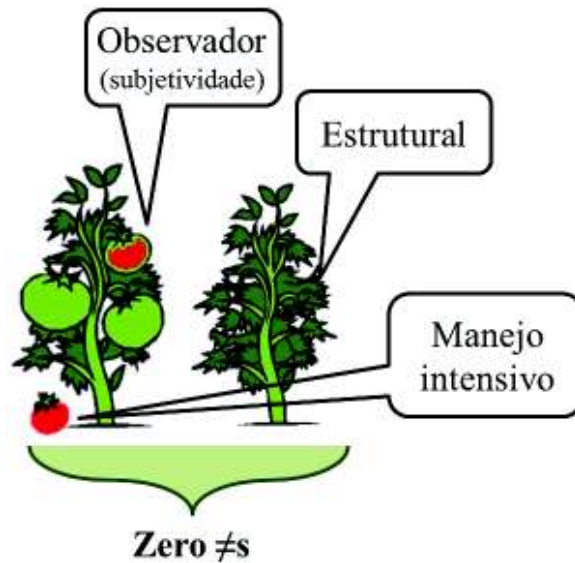
2.1 ANÁLISE ESTATÍSTICA EM OLERÍCOLAS

As olerícolas são culturas agrícolas com importância em nível nacional e global, uma vez que são utilizadas quase que exclusivamente na alimentação humana. Devido a essa importância, existe uma demanda crescente por informações que devem ser embasadas em resultados de experimentos.

A qualidade das informações repassadas aos agricultores dependerá muito da precisão destes resultados. Porém, a precisão experimental em experimentos com olerícolas é dificultada devido a aspectos inerentes a cultura, como a heterogeneidade entre plantas e entre colheitas, a maturação desuniforme das plantas e o ponto subjetivo da colheita (LÚCIO; SARI, 2017). Além disso, em olerícolas com múltiplas colheitas, a mensuração das variáveis é realizada repetidamente em uma mesma planta ou parcela, caracterizando-as como medidas repetidas no tempo.

Outra característica que causa elevada variabilidade é a inflação de zeros no banco de dados. No caso de olerícolas, além do zero estrutural (quando a planta realmente não possui frutos aptos a serem colhidos), existem outros “tipos de zeros”. Um é oriundo da subjetividade do ponto de colheita. Como os frutos são colhidos após a mudança de coloração, para um avaliador determinado fruto pode estar apto a ser colhido, e para outro não. Por fim, outra fonte de zeros é oriunda de quedas inevitáveis de frutos das plantas, devido a manejo intensivo das culturas (Figura 1).

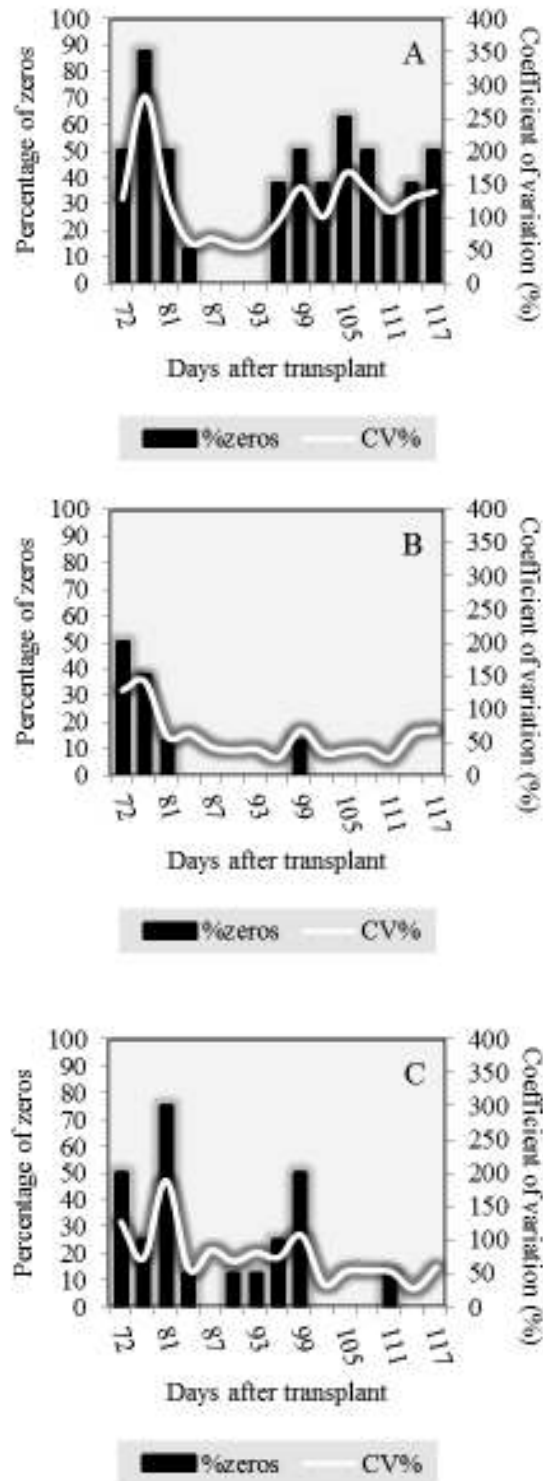
Figura 1 - Diferentes tipos de zeros observados em culturas olerícolas de múltiplas colheitas.



Para que um experimento apresente resultados consistentes é necessário um planejamento subsidiado por técnicas que minimizem o erro experimental e por procedimentos estatísticos adequados. No caso das olerícolas de múltiplas colheitas, a minimização do erro experimental é dificultada pela elevada variabilidade entre as plantas, entre as colheitas (são intrínsecas a cultura, e não podem ser reduzidas) (Figura 2) e a correlação entre as observações devido às medições sistemáticas (múltiplas colheitas) em uma mesma planta ou parcela. Também existe a dificuldade de determinar um modelo da análise de variância (ANOVA) que avalie todas estas fontes de variação, dificultando a análise do experimento.

No caso da ANOVA, a superdispersão dos dados e o excesso de zero (causada pela variabilidade entre plantas e colheitas) levam a violação da homocedasticidade e normalidade dos resíduos (LÚCIO et al., 2012). Além disso, ao declarar as colheitas (medidas repetidas) como um fator fixo no modelo, a pressuposição da independência dos resíduos é violada, pois as colheitas não podem ser casualizadas nas parcelas (FERNANDEZ, 2007; LÚCIO et al., 2012; LÚCIO; SARI, 2017). Em experimentos com medidas repetidas, regressões lineares e não lineares podem ser utilizadas com ferramenta de análise alternativa (CASELLA, 2008).

Figura 2 - Porcentagens de zero e coeficiente de variação (CV%) entre plantas por colheita para a variável número de frutos por planta nos genótipos de tomate Gaúcho (A), Cordillera (B) e Janaína (C). Fonte: dados pessoais.



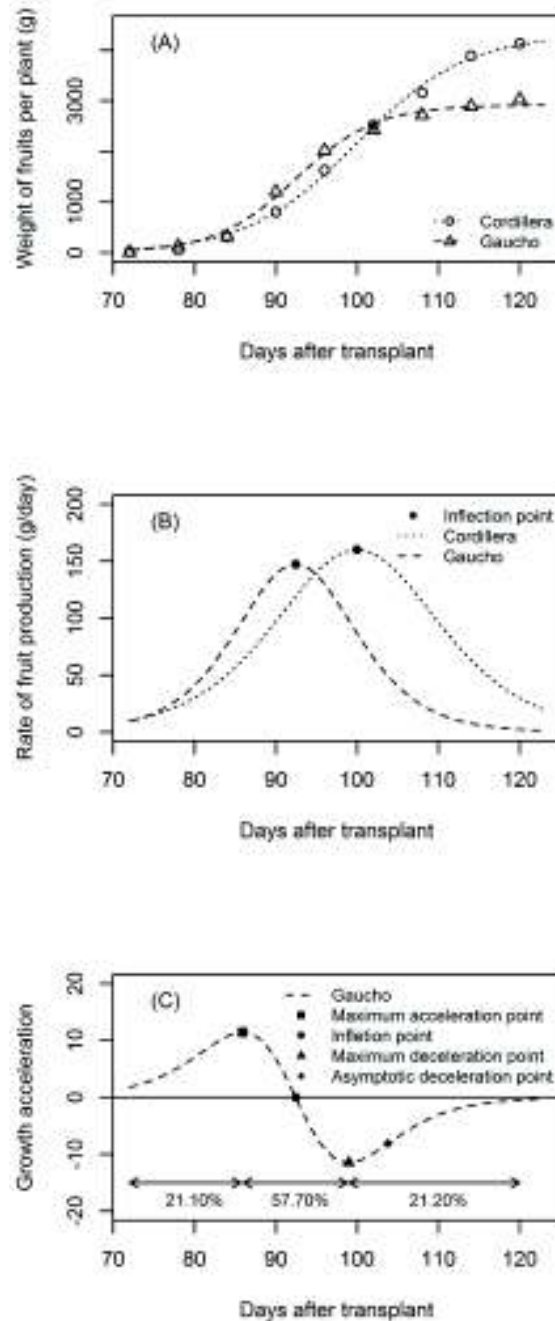
As dificuldades relatadas levam o pesquisador a optar pelo agrupamento das plantas (utilizando parcelas compostas por muitas plantas) e das colheitas. O agrupamento de plantas reduz a presença de zeros no banco de dados, enquanto que o agrupamento das colheitas reduz a probabilidade de violação dos pressupostos de homocedasticidade e independência dos resíduos (LÚCIO et al., 2012; LÚCIO; SARI, 2017). Por isso os pesquisadores tendem a analisar os experimentos em olerícolas desconsiderando as observações realizadas em cada colheita, analisando apenas a produção total. Porém isso leva a uma perda de informações sobre o comportamento produtivo das culturas ao longo das colheitas.

A ANOVA, usualmente utilizada na experimentação agrícola, foi desenvolvida para dados com distribuição gaussiana. Porém quando os dados que não seguem a distribuição normal, principalmente quando inflacionados com zero (caso das olerícolas), os resíduos também não seguirão. Com o desenvolvimento computacional, as análises estatísticas baseadas em modelos da ANOVA podem ser substituídas por métodos estatísticos mais modernos, como modelos mistos, modelos lineares generalizados e por modelos mistos não lineares (FERNANDEZ, 2007). Esses métodos eram pouco utilizados devido ao grande esforço computacional necessário para realizar cálculos, barreira que foi transposta pela evolução computacional e de softwares estatísticos.

A produção acumulada de olerícolas de múltiplas colheitas tende a ter um comportamento sigmoide (Figura 3A) e, por isso, a utilização de regressões não lineares pode ser uma alternativa de análise para este tipo de experimento. Entre as regressões não lineares, os modelos de crescimento são uma ótima alternativa para descrever comportamentos sigmóides (SEBER; WILD, 2003). Os modelos fornecem estimativas de seus parâmetros com interpretação biológica que podem ser utilizadas como variáveis explicativas do comportamento produtivo das olerícolas (LÚCIO et al., 2015; LÚCIO et al., 2016a; LÚCIO et al., 2016b).

A estimação dos parâmetros depende do quão próximo ao comportamento linear é o modelo não linear, uma vez que a estimação é realizada pelo método dos mínimos quadrados (BATES; WATTS, 1988; DRAPER; SMITH, 1998; KUTNER et al., 2004). Como é o comportamento linear dos modelos que permite estimar os parâmetros pelo método dos mínimos quadrados, o conhecimento do grau de não linearidade é fundamental para que o pesquisador possa utilizá-los em suas análises estatísticas. Esses aspectos serão abordados de modo a orientar o pesquisador a utilizar adequadamente os modelos de crescimento como ferramenta de análise estatística em olerícolas.

Figura 3 - Ajuste do modelo logístico (A), taxa de produção de frutos (B) e taxa de aceleração (C) dos genótipos Cordillera e Gaucho Fonte: dados pessoais.



2.2 ESTIMAÇÃO DOS PARÂMETROS EM MODELOS NÃO LINEARES

Em modelos lineares os efeitos dos parâmetros aparecem de forma aditiva no modelo, e por isso eles são considerados lineares. Já nos modelos não lineares pelo menos um dos parâmetros aparece de forma não aditiva.

$$E(y) = m + b_j + t_i \quad (1)$$

$$E(y) = \frac{\hat{\beta}_1}{1 + e^{(\hat{\beta}_2 - \hat{\beta}_3 x_i)}} \quad (2)$$

Os modelos (1) e (2) são representações de um modelo linear e não linear, respectivamente. O primeiro (1) representa um modelo da análise de variância (ANOVA) de um experimento no delineamento blocos ao acaso. Os efeitos de bloco (b_j) e tratamentos (t_i) aparecem de forma aditiva no modelo, e seus efeitos são desvios em relação a média geral (m). O segundo modelo (2) é o logístico, onde o parâmetro β_1 é o numerador de uma razão e os parâmetros β_2 e β_3 são o expoente da base e (número de Euler), ou seja, seu efeito é exponencial. Percebe-se facilmente que o efeito dos parâmetros em (2) não é aditivo como em (1).

A forma de estimar os parâmetros é afetada dependendo de como eles aparecem no modelo. Em modelos não lineares a estimação é realizada minimizando a soma de quadrados dos resíduos, que é a diferença entre o valor observado e o valor estimado pelo modelo. Este método é conhecido como o “método dos mínimos quadrados”. Ao derivar a soma de quadrados dos resíduos em relação aos parâmetros é obtido o sistema de equações normais. Esse sistema pode ser resolvido analiticamente em modelos lineares, e os parâmetros facilmente estimados por (DRAPER; SMITH, 1998; KUTNER et al., 2004):

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (3)$$

Em (3), \mathbf{X} é uma matriz de dimensões $n \times p$ conhecida como matriz delineamento e \mathbf{y} e $\hat{\boldsymbol{\beta}}$ são os vetor de observações e parâmetros de tamanhos $n \times 1$ e $p \times 1$, respectivamente. A solução analítica do sistema de equações e, conseqüentemente, a obtenção das estimativas dos parâmetros em modelos lineares é possível pois \mathbf{X} não dependem dos parâmetros. No caso não linear as derivadas dependerão de ao menos um dos parâmetros do modelo, e a utilização de métodos iterativos passa a ser necessária para resolver o sistema de equações normais não linear (BATES; WATTS, 1988; SEBER; WILD, 2003).

2.2.1 Métodos iterativos

A estimação dos parâmetros dos modelos não lineares também é realizada via método dos mínimos quadrados. Porém a resolução do sistema de equações normais é realizada utilizando aproximações lineares da função (modelo). A expansão de Taylor de primeira ordem de uma função é dada por

$$f(x, \boldsymbol{\theta}) = f(x, \boldsymbol{\theta}^0) + \frac{\partial f(x, \boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \quad (4)$$

Na expressão (4), $\boldsymbol{\theta}$ são os parâmetros do modelo e $\boldsymbol{\theta}^0$ são valores iniciais dos parâmetros determinados arbitrariamente. A expressão (4) pode ser simplificada por $f(\boldsymbol{\theta}) = f(\boldsymbol{\theta}^0) + \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^0)$, onde \mathbf{F} é a matriz $n \times p$ de derivadas de primeira ordem. Minimizando a soma de quadrados dos resíduos, tem-se

$$\begin{aligned} S(\boldsymbol{\theta}) &= \sum_{i=1}^n (y - f(\hat{\boldsymbol{\theta}}))^2 \\ S(\boldsymbol{\theta}) &= \sum_{i=1}^n (y - f(\boldsymbol{\theta}^0) - \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^0))^2 \end{aligned} \quad (5)$$

Substituindo $y - f(\boldsymbol{\theta}^0)$ por ε , temos $S(\boldsymbol{\theta}) = \sum_{i=1}^n (\varepsilon - \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^0))^2$, cuja solução (por analogia ao método de mínimos quadrados do modelo linear) é

$$\boldsymbol{\theta} - \boldsymbol{\theta}^0 = (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}'\boldsymbol{\varepsilon} \quad (6)$$

Em (6) \mathbf{F} é a matriz $n \times p$ de derivadas parciais (jacobiana) $\frac{\partial f(x, \boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^0}$. De (6) obtemos o algoritmo de Gauss-Newton, dado por

$$\hat{\boldsymbol{\theta}}^* = \boldsymbol{\theta}^0 + (\mathbf{F}^0{}'\mathbf{F}^0)^{-1} \mathbf{F}^0{}'\boldsymbol{\varepsilon} \quad (7)$$

Para iniciar o processo iterativo é necessário dar um valor inicial arbitrário aos parâmetros. Em (7), θ^0 corresponde a este valor inicial quando é iniciado o algoritmo, ou ao valor do passo imediatamente anterior. Os valores de $\hat{\theta}^*$ obtidos em cada passo do algoritmo se aproxima de θ , até a convergência (BATES; WATTS, 1988; DRAPER; SMITH, 1998; SEBER; WILD, 2003). Ressalta-se que o algoritmo de Gauss-Newton é amplamente utilizado em softwares estatísticos, como no R e no SAS.

2.2.2 Valores iniciais dos parâmetros

Os valores iniciais dos parâmetros são importantes para que o método iterativo alcance a convergência (RITZ; STREIBIG, 2008). Quanto mais próximo de $\hat{\theta}$ for o valor inicial, mais rápida é a convergência (menor será o número de iterações). No momento de obter esses valores o pesquisador poderá adotar as seguintes opções (BATES; WATTS, 1988; ARCHONTOULIS; MIGUEZ, 2015):

- a) Usar informações encontradas na literatura;
- b) Método gráfico: valores iniciais e, posteriormente, mudanças no valor dos parâmetros são realizadas arbitrariamente no modelo. Quando a curva se aproxima dos valores aproximados, têm-se os valores iniciais.
- c) Interpretar o comportamento das derivadas do modelo.
- c) Linearizar os modelos (se possível): modelos que são intrinsicamente lineares podem ser linearizados. Os parâmetros do modelo que foi linearizado podem ser estimados e utilizados como valores iniciais.

2.2.3 Medidas de não linearidade

O algoritmo de Gauss-Newton é baseado em uma aproximação linear de Taylor de primeira ordem, como descrito em (4). Quando a função não linear é aproximadamente linear, as estimativas dos parâmetros passam a ter características mais próximas daquelas observadas em estimativas dos parâmetros de um modelo linear. Portanto, medidas de não linearidade são importantes, pois possibilitam ao pesquisador verificar o quão próximo do linear é o comportamento de um modelo não linear. Quando o comportamento não é aproximadamente linear, as estimativas dos parâmetros são (muito) viesadas, os intervalos de confiança não são

precisamente estimados e as hipóteses (quanto aos parâmetros) não podem ser testadas (BATES; WATTS, 1988; SEBER; WILD, 2003; RITZ; STREIBIG, 2008)

Bates e Watts (1988) utilizaram técnicas de geometria diferencial para determinar duas medidas de curvatura que medem o grau de não linearidade do modelo: a não linearidade intrínseca, que é inerente ao modelo, e a não linearidade paramétrica, que depende da parametrização (BATES; WATTS, 1988). A função *rms.curv* do pacote *MASS* no software R (VENABLES; RIPLEY, 2002; RIPLEY, 2017) retorna os seguintes valores

$$\begin{aligned} c^I & \sqrt{F_{(\alpha;p,n-p)}} \\ c^\theta & \sqrt{F_{(\alpha;p,n-p)}} \end{aligned} \quad (8)$$

Em (8), c^I mede a não linearidade intrínseca; c^θ mede a não linearidade paramétrica; p é o número de parâmetros do modelo; n é o número de observações; e $F_{(\alpha;p,n-p)}$ é o quantil (α) da distribuição F com p e $n-p$ graus de liberdade. Quanto menor esses valores, maior é a aproximação linear da função (BATES; WATTS, 1988; SEBER; WILD, 2003).

Valores de $c^I F_{(\alpha;p,n-p)}$ menores que 0.3 e de $c^\theta F_{(\alpha;p,n-p)}$ menores que 1 indicam uma boa aproximação linear. Vale ressaltar que esses valores de referencia são discutíveis. Porém é aconselhado que os valores baixos destas estatísticas devem ser considerados na hora de selecionar o modelo, pois eles indicam que os parâmetros estimados são aproximadamente não viesados. Infelizmente esse tipo de procedimento é pouco utilizado e a maioria das situações a seleção dos modelos é realizada apenas pela qualidade de ajuste (via R^2 , AIC, etc.) e não pela qualidade das estimativas dos parâmetros.

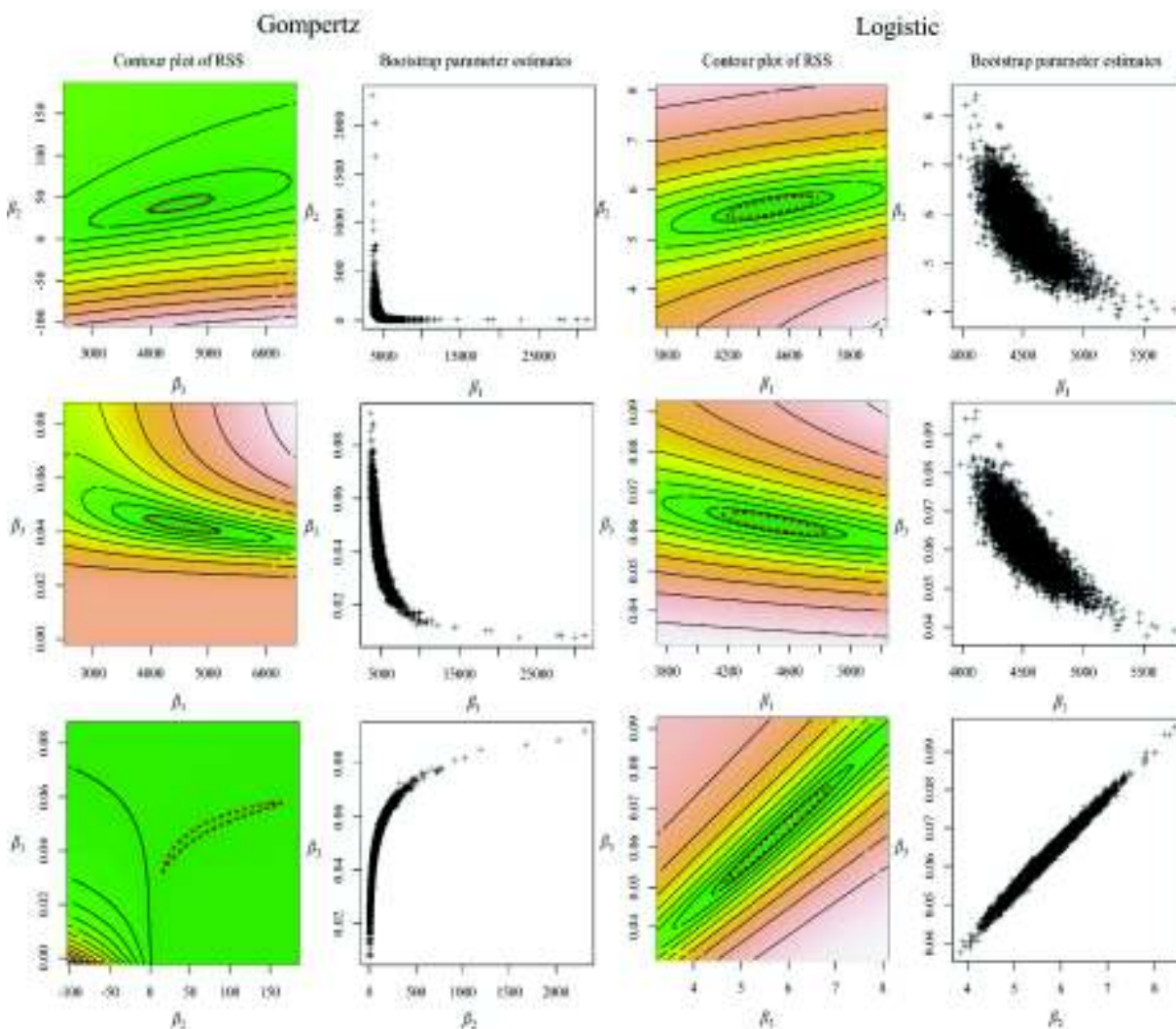
2.2.4 Gráficos de contornos

Uma forma usual de visualizar a aproximação linear dos modelos é via contornos da soma de quadrados dos resíduos (SQ_{Erro}). Os contornos também permitem verificar se o processo iterativo convergiu para mínimos locais e levaram a estimação de parâmetros inadequados (BATY et al., 2015). Esses contornos correspondem aos valores dos parâmetros que resultam em uma SQ_{Erro} qualquer. À medida que o processo iterativo avança, o contorno (SQ_{Erro}) estimado passa a ficar próximo do contorno verdadeiro e $\hat{\theta}$ passa a ficar próximo de θ (convergência). Quando um modelo não linear tem um comportamento próximo ao linear,

as estimativas dos parâmetros do modelo não linear têm propriedades próximas às estimativas de um modelo linear (CARVALHO et al., 2010).

Contornos com forma elíptica indicam que o modelo é aproximadamente linear. Já as elipses curvadas, chamadas de “banana-shaped” (forma de banana) (Figura 4) por Seber e Wild (2003), indicam problema de aproximação linear dos modelos, prejudicando as estimativas dos parâmetros (viesados) e dos intervalos de confiança, causando problemas inferenciais (inviabilizam os testes de hipóteses) (SEBER; WILD, 2003).

Figura 4 - Contornos da soma de quadrados dos resíduos e estimativas dos parâmetros obtidos por 5000 reamostragens bootstrap do modelo Gompertz e logístico ajustado a dados de produção de berinjela. Fonte: dados pessoais.



Percebe-se na Figura 4 o problema em ajustar modelos sem tomar o devido cuidado com a não linearidade. No modelo logístico a região de confiança de 95% de Beale (1960)

(contorno em vermelho) coincide com os contornos, que são elipsoides. No modelo de Gompertz os contornos apresentaram a forma de banana, ou a região de confiança de Beale não coincidem com o contorno das SQ_{Erro} , indicando problemas de aproximação linear do modelo. Mesmo com elevada qualidade de ajuste, problemas inferenciais graves são observados, inviabilizando a utilização do modelo de Gompertz. Apesar de possuir correlação entre dois parâmetros (β_2 e β_3) o modelo logístico deve ser selecionado pois possui contornos na forma elíptica e convergiu para um mínimo global com $\hat{\theta}$ próximo a θ .

2.2.5 Bias

A aproximação linear dos modelos não lineares é que possibilitam a utilização dos métodos dos mínimos quadrados para estimar os parâmetros em modelos não lineares. Assintoticamente, as estimativas dos parâmetros não lineares estimados por mínimos quadrados são normalmente distribuídos, não viesados e com variância dada por $\sigma^2 (\mathbf{F}'\mathbf{F})^{-1}$. Porém, quando o n é moderado ou pequeno a tendência é que as estimativas possuam um viés e a variância exceda $\sigma^2 (\mathbf{F}'\mathbf{F})^{-1}$. O viés das estimativas é consequência do quão próximo é linear são os modelos não lineares (Seber and Wild, 2003).

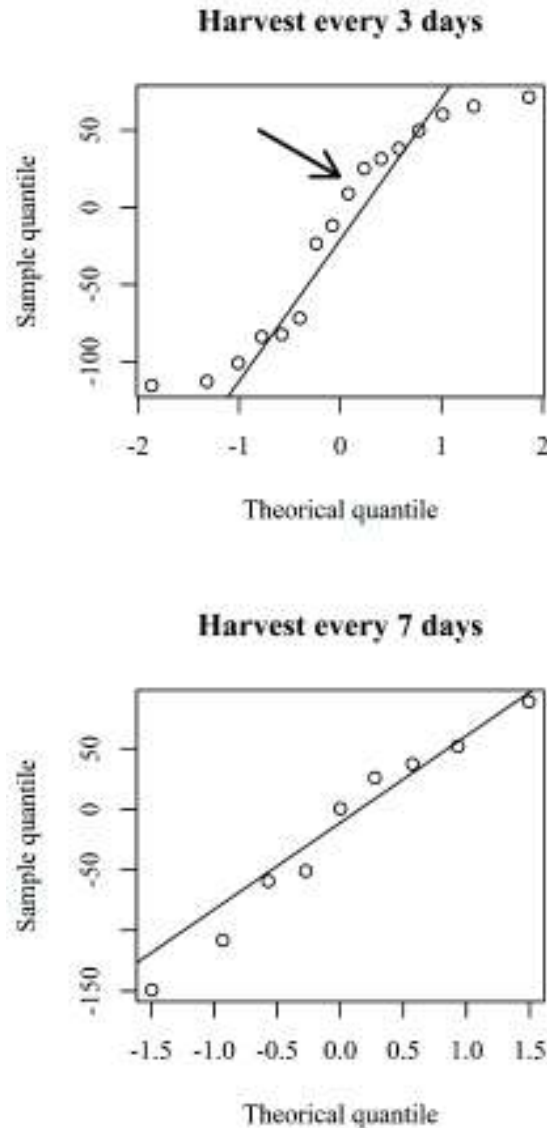
Existem inúmeras formas de verificar o BIAS das estimativas dos parâmetros através do a) vício de Box (1971), b) pela diferença entre a média das estimativas obtidas por reamostragem “*bootstrap*” (>1000) e a estimativa obtida pelo método de mínimos quadrados e c) pela verificação da normalidade da distribuição empírica das estimativas “*bootstrap*”. Vícios elevados (>1%) e não normalidade da distribuição empírica indica falta de aproximação linear (ZEVIANI et al., 2012).

2.3 PRESSUPOSTOS DOS MODELOS

Os pressupostos de um modelo não linear são a homocedasticidade, independência e normalidade dos resíduos. O cumprimento desses pressupostos não é necessário para que os parâmetros sejam estimados pelo método dos mínimos quadrados. Porém, o não cumprimento afeta a precisão da estimativa dos intervalos de confiança e compromete os testes de hipóteses. Por isso o cumprimento destes pressupostos é necessário para que as inferências sejam corretas e as conclusões baseadas no modelo sejam confiáveis (RITZ; STREIBIG, 2008).

Em olerícolas a heterogeneidade e a independência dos resíduos são pressupostos que podem ser facilmente violados no momento de ajustar um modelo (LÚCIO et al, 2015). Isso é observado pois a variabilidade de produção entre as plantas tende a ser grande no início das colheitas, gerando heterocedasticidade. Já a dependência dos resíduos ocorre pois as colheitas se constituem como medidas repetidas. Quando a intensidade de colheitas é elevada (realizadas em períodos curtos), esse problema tende a se agravar. Isso ocorre pois valores repetidos (mesmo número ou massa de frutos) ou muito próximos são mensurados consecutivamente em uma mesma planta ou parcela, uma vez que a planta não foi capaz de produzir (ou produziu poucos) frutos aptos a serem colhidos. Como consequência, os erros tornam-se correlacionados uma vez que possuem um mesmo sinal ao longo de várias avaliações (SEBER; WILD, 2003). Esse comportamento é facilmente percebido em um gráfico quantil-quantil (Figura 5).

Figura 5 - QQ-Plot dos resíduos do modelo logístico ajustado a dados de produção (massa de frutos) do genótipo de tomate Cordillera, colhido a cada três e sete dias. Fonte: dados pessoais.



*A seta indica uma sequencia de resíduos com um mesmo sinal quando as colheitas são realizadas mais intensamente.

2.3.1 Diagnósticos

Existem dois tipos de procedimentos que podem ser utilizados para verificar o atendimento ou não as pressuposições dos modelos: testes estatísticos e análise gráfica. Os testes estatísticos tem a vantagem de serem facilmente interpretados. As estatísticas dos testes são utilizadas para rejeitar ou não a hipótese nula H_0 , que afirma que os erros são normais, homocedásticos ou independentes, dependendo do teste que se está sendo utilizando. Então, não rejeitando H_0 , concluímos que os pressupostos foram cumpridos. Geralmente os

pesquisadores utilizam o *p-valor* para verificar o cumprimento dos pressupostos (quando *p-valor* > 0,05, a hipótese H_0 não é rejeitada). Os testes mais utilizados estão identificados na Tabela 1.

Tabela 1 - Testes estatísticos recomendados para testar os pressupostos e suas respectivas funções no software R.

Pressuposto	Testes	Função R	Pacote R
Homocedasticidade	Levene	levene.test()	nlrwr ¹
	Bartlett	bartlett.test()	stats ²
	Breusch–Pagan	bptest()	car ³
Normalidade	Shapiro-Wilk	shapiro.test()	stats
	Kolmogorov-Smirnov	lillie.test()	nortest ⁴
Independência	Durbin-Watson	durbinWatsonTest()	car
		dwtest	lmtest ⁵

¹ RITZ; STREIBIG (2008), RITZ, 2013; ² R CORE TEAM (2017); ³FOX; WEISBERG (2011);⁴ GROSS; LIGGES (2015);⁵ HOTHORN et al. (2017).

Percebe-se que há uma facilidade de interpretação dos testes estatísticos. Porém eles não mostram ao pesquisador a causa da violação dos pressupostos. Por isso a análise gráfica dos resíduos deve ser, também, realizada. A homocedasticidade pode ser verificada através de gráficos de resíduos *vs* colheitas (Figura 6). A normalidade dos resíduos pode ser verificada através de histogramas, box-plot e QQ-Plot (Figura 7). A interpretação dos gráficos é realizada da seguinte forma (LÚCIO; SARI, 2017):

a) Histograma de frequência: Valores concentrados nos extremos do gráfico indicam distribuição assimétrica e não normal. Se a assimetria é negativa, os valores estão concentrados a direita (Figura 7E e 7F) e se a assimetria é positiva, os valores estão concentrados a esquerda (Figura 7C e 7D). Já a curtose é uma medida que caracteriza o achatamento da curva, sendo ela é denominada de leptocúrtica quando é “pontaguda” (Figura 7G e 7H), e platicúrtica quando é muito “achatada” (Figura 7I e 7J). Ambas as condições indicam ausência de normalidade (Figura 7A e 7B);

b) QQ-plot: valores observados em forma de “arco” (Figura 7C, 7D, 7E e 7F) indicam distribuição assimétrica; valores observados sob a linha indicam distribuição normal ou seja, distribuição simétrica e mesocúrtica (Figura 7A e 7B); valores em forma de “S” indicam que a curva é leptocúrtica ou platicúrtica (Figura 7G, 7H, 7I e 7J);

c) Box-plot: caixa “deslocada” para a parte superior do gráfico indica assimetria negativa (Figura 7C e 7D); caixa “deslocada” para a parte inferior do gráfico indica assimetria

positiva (Figura 7E e 7F); caixa “centralizada” indica distribuição normal (Figura 7A e 7B); caixa “estreita” indica curva leptocúrtica (Figura 7G e 7H); caixa “longa” indica curva platicúrtica (Figura 7I e 7J).

Figura 6 - Gráfico dos resíduos (simulados) vs colheitas indicando heterocedasticidade e homocedasticidade dos resíduos. Fonte: dados pessoais.

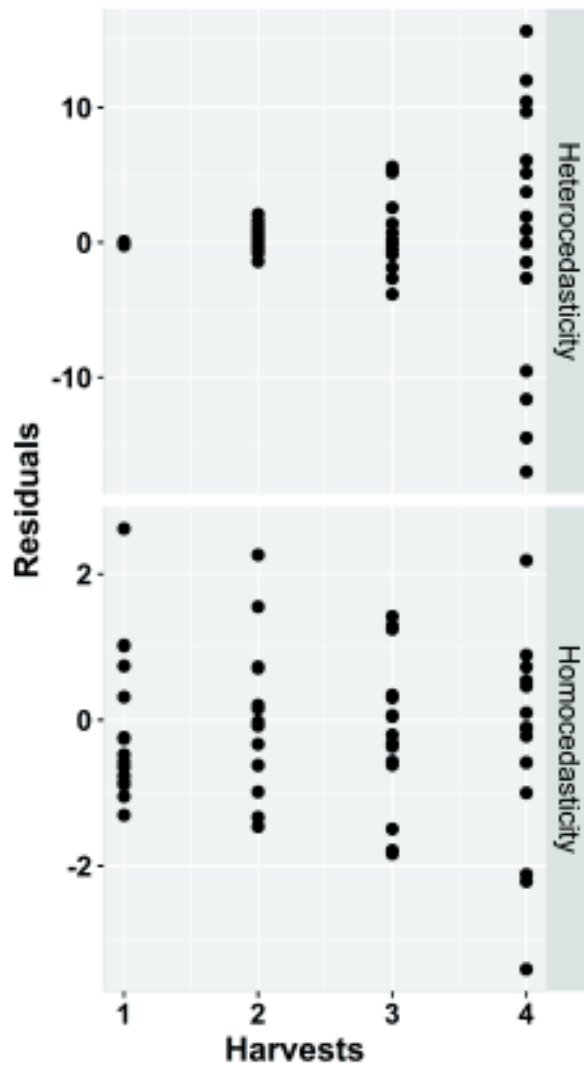
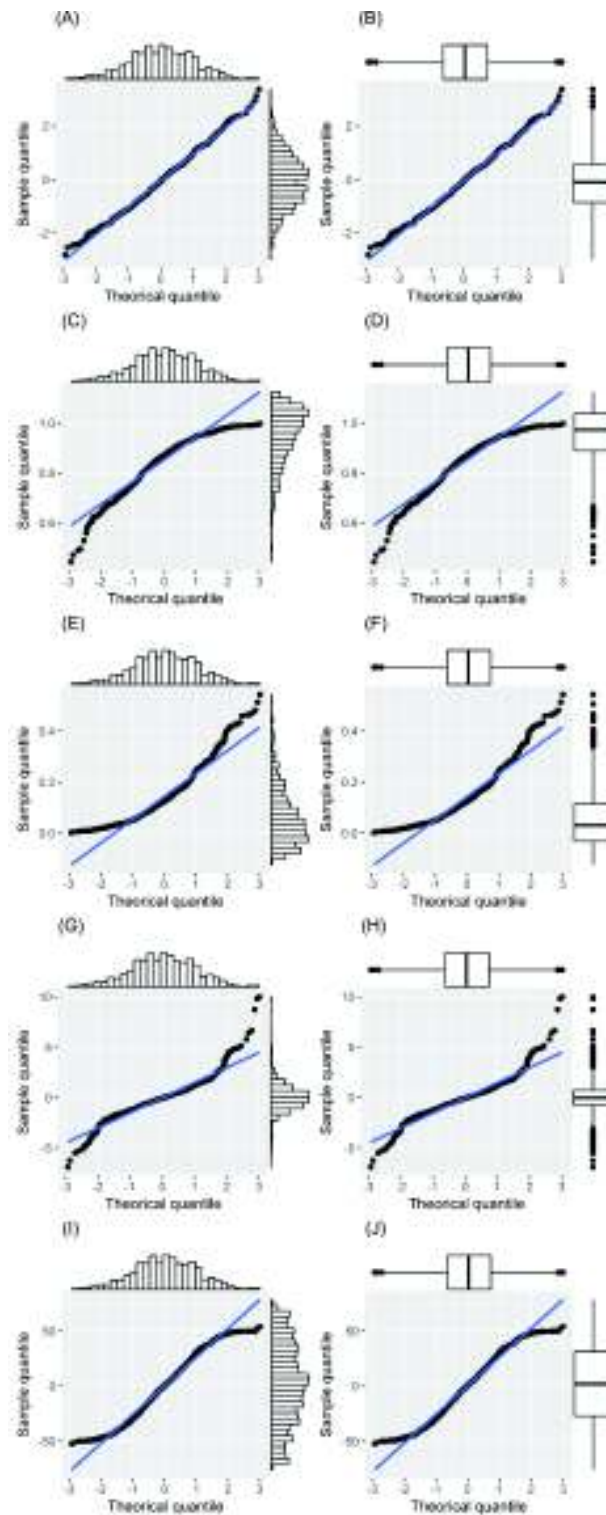
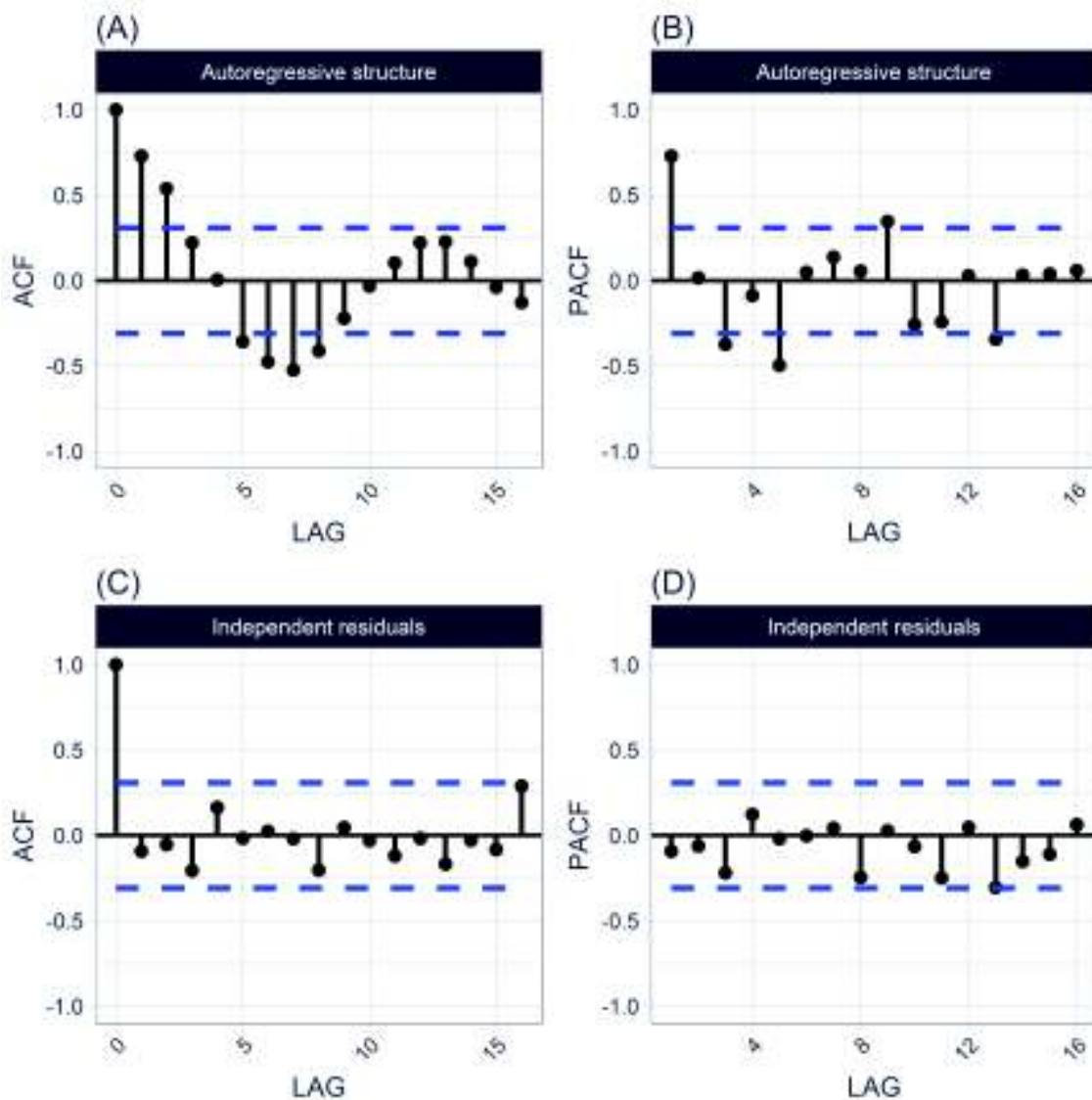


Figura 7 - Gráfico quantil-quantil, histogramas de frequência e box-plot em condições de normalidade (A,B), assimetria a esquerda (C,D), assimetria a direita (E,F), distribuição leptocúrtica (G,H) e distribuição platicúrtica (I,J), utilizando dados simulados.



O diagnóstico de independência dos resíduos pode ser realizado através da utilização de gráficos de autocorrelação e autocorrelação parcial. A partir de comportamento do gráfico de autocorrelação é possível determinar a estrutura de correlação dos resíduos. Quando os resíduos são independentes, a autocorrelação entre os resíduos fica dentro do limite imposto pelo erro padrão da autocorrelação nula, obtida por $n^{-1/2}$ (SEBER; WILD, 2003), onde n é o tamanho da amostra (Figura 8C e 8D).

Figura 8 - Gráficos de autocorrelação (ACF) e autocorrelação parcial (PAF) obtidos a partir de resíduos correlacionados (A,B) e independentes (C,D) do modelo logístico ajustado a dados de produção de frutos de morangueiro. Fonte: dados pessoais.



*A linha pontilhada em azul corresponde ao erro padrão quando a autocorrelação é nula, obtido por $n^{-1/2}$.

No caso das olerícolas é comum que os erros correlacionados tenham uma estrutura autoregressiva (AR). Essa estrutura é caracterizada pela queda exponencial ou oscilatória da autocorrelação residual (Figura 8A). O grau p de AR é determinado graficamente pelo gráfico de autocorrelação parcial (Figura 8B) (BATES; WATTS, 1988). Seber e Wild (2003) descrevem o comportamento dos erros correlacionados em estruturas de médias móveis (MA) e médias móveis e autoregressivos (ARMA).

2.3.2 Contornando o problema de violação dos pressupostos

Como dito anteriormente, as pressuposições de normalidade, homogeneidade e independência dos resíduos não é necessária para estimar os parâmetros pelo método dos mínimos quadrados. Porém, quando há heterocedasticidade ou dependência nos resíduos, os parâmetros estimados são ineficientes (SEBER; WILD, 2003; RITZ; STREIBIG, 2008; MUNIZ et al., 2017). Quando esses pressupostos são cumpridos, os parâmetros são estimados de forma mais precisa e o grau de ajustamento do modelo é melhor (SAVIAN; MUNIZ, 2007; SOUSA et al., 2014; MUNIZ et al., 2017).

No caso de olerícolas, a heterogeneidade das plantas e as medidas repetidas geram problemas de heterocedasticidade e, conseqüentemente, de não normalidade e dependência dos resíduos. Uma forma prática de reduzir o problema de heterocedasticidade é agrupar colheitas, ou utilizar parcelas grandes (11 plantas para número de frutos e quatro plantas para massa no caso do tomate) no caso de experimentos delineados (LÚCIO et al., 2016c; LÚCIO; SARI, 2017). A desvantagem nesse caso é a perda de eficiência no uso da área experimental devido a necessidade de parcelas muito grandes gerando uma redução no número de repetições e/ou tratamentos do experimento. No caso da independência dos resíduos, espaçar mais as colheitas é uma ótima alternativa (Figura 5).

Apesar das opções relatadas acima, os problemas de heterocedasticidade e dependência dos resíduos podem ser solucionados pela estimação dos parâmetros utilizando os métodos de mínimos quadrados ponderados e mínimos quadrados generalizados, respectivamente. O procedimento pelo método de mínimos quadrados ponderados é o mesmo do método de mínimos quadrados ordinários, porém a função variância é ponderada por um peso. Já pelo método de mínimos quadrados generalizados é dada aos resíduos uma estrutura de correlação (por exemplo, uma estrutura autoregressiva) via um parâmetro ϕ desconhecido. Esses procedimentos são facilmente implementados no software R pelo pacote *gnls* do pacote *nlme* (PINHEIRO et al., 2017).

2.4 SELEÇÃO DE MODELOS

Ao explicar um fenômeno por um modelo, sempre haverá erros. Porém, alguns modelos são mais úteis que os outros e uma das formas de selecionar um modelo é quantificar esse “erro”. Uma das formas mais comuns de verificar a "bondade do ajuste" é através dos critérios de informação. Esses critérios se baseiam na magnitude da função de verossimilhança $L(\hat{\theta})$, que é o BIAS (erro) da função $f(x, \theta)$, geralmente desconhecido (EMILIANO et al., 2014). Os três critérios mais utilizados serão apresentados a seguir:

a) Critério de informação de Akaike (*AIC*): proposto por Akaike (1974), este critério é obtido por:

$$AIC = -2L(\hat{\theta}) + 2p \quad (9)$$

Em (9), $L(\hat{\theta})$ é o valor máximo da função de verossimilhança e p é o número de parâmetros do modelo.

b) Critério de informação de Akaike corrigido (*AICc*): A derivada de segunda ordem do AIC deu origem ao *AICc*, proposto por Sugiura (1978):

$$AICc = -2 \log L(\hat{\theta}) + 2p \left(\frac{n}{n-p-1} \right) \quad (10)$$

Em (10), $L(\hat{\theta})$ é o valor máximo da função de verossimilhança, p é o número de parâmetros do modelo e n é o número de observações. O *AICc* deve ser utilizado quando $\frac{n}{p}$ é baixo, geralmente menor que 40 (Emiliano et al., 2014).

c) Critério de informação Bayesiano (*BIC*): foi proposto por Schwarz (1978), sendo obtido por

$$BIC = -2L(\hat{\theta}) + p \log n \quad (11)$$

Em (11), $L(\hat{\theta})$ é o valor máximo da função de verossimilhança, p é o número de parâmetros do modelo e n é o número de observações. Essas estatísticas são utilizadas quando a estimação dos parâmetros é realizada pelo método de máxima verossimilhança. Os critérios acima descritos são utilizados para selecionar um dentre vários modelos. Quanto menor os valores de AIC, AICc e BIC maior a qualidade do ajuste e, conseqüentemente, melhor é o modelo. Emiliano et al. (2014) recomenda os critérios de AIC e AICc para selecionar modelos de crescimento com tamanhos de amostra entre 20 e 30.

Os valores dos critérios de informação são utilizados para casos em que os modelos não são aninhados. No caso de modelos aninhados, podem-se utilizar as estatísticas qui-quadrado (χ^2) e F para selecionar os modelos (RITZ; STREIBIG, 2008; ARCHONTOULIS; MIGUEZ, 2015). Estas estatísticas mostram se o incremento no ajustamento através da adição de mais um parâmetro é ou não significativo. Detalhes sobre o teste F e qui-quadrado serão apresentados na seção 2.5.2.

2.4.1 Qualidade do ajustamento

A qualidade do ajuste atesta a capacidade do modelo de predizer os valores observados. Essas estatísticas também podem ser usadas para selecionar modelos. As estatísticas mais utilizadas são o coeficiente de determinação (R^2), o coeficiente de determinação ajustado (R_{aj}^2), o desvio padrão residual (EPA) e a eficiência de modelagem (ME). Estas estatísticas são calculadas da seguinte forma:

$$R^2 = 1 - \frac{SQ_{Erro}}{SQ_{Total}}$$

$$R_{aj}^2 = \left(R^2 - \frac{p}{n-1} \right) \left(\frac{n-1}{n-p-1} \right)$$

$$EPA = \sqrt{QM_{Erro}}$$

$$ME = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}$$
(12)

Em (12), SQ_{Erro} é a soma de quadrados dos resíduos, SQ_{Total} é a soma de quadrados totais, QM_{Erro} é o quadrado médio do erro, p é o número de parâmetros, n é o número de observações, Y_i são os valores observados, \hat{Y}_i são os valores estimados e \bar{Y}_i é a média das observações na i -ésima colheita (quando há repetições).

Quanto mais próximos de um os valores de R^2 e ME , melhor a qualidade de ajustamento do modelo. Modelos com maior valor de R^2_{aj} devem ser selecionados. Por fim, quanto menor o valor do EPA , melhor a qualidade do ajuste. O R^2 pode ser utilizado quando os modelos tem o mesmo número de parâmetros e observações. É importante ressaltar que a expressão apresentada para o coeficiente de determinação é adequada apenas para modelos com intercepto. Quando o modelo possui intercepto, o valor de R^2 corresponde ao valor do coeficiente de correlação entre os valores observados e estimados ao quadrado (r^2_{yx}). No caso de dúvidas quanto a presença ou não do intercepto, recomenda-se utilizar como medida o valor de r^2_{yx} .

2.4.2 Como selecionar o modelo?

A seleção de modelos de crescimento não pode ser realizada apenas em função da qualidade de ajuste aos dados. Para que as estimativas dos parâmetros com interpretação biológica possam ser usadas para explicar determinado fenômeno biológico (SEBER; WILD, 2003), a precisão da sua estimação deve ser levada em consideração. No caso de modelos não lineares, essa avaliação é realizada com base na não linearidade intrínseca e paramétrica (BATES; WATTS, 1988). Em casos de não linearidade elevada, os estimadores são viesados e os intervalos dos parâmetros não são precisos, prejudicando a interpretação dos resultados.

Portanto, a seleção dos modelos não lineares deve levar em consideração a precisão de estimação dos parâmetros. Valores que mensurem a capacidade de predição, como os relatados anteriormente, também podem, e devem, ser utilizados, desde que para selecionar modelos cujos parâmetros foram estimados de forma precisa. Sem levar em consideração este aspecto, o pesquisador pode selecionar de forma equivocada os modelos, muitas vezes com parâmetros que não condizem com a realidade.

2.5 MODELOS DE CRESCIMENTO

Os modelos de crescimento são funções empíricas flexíveis utilizadas para descrever o crescimento de alguma medida (largura, altura, massa, etc) ao longo do tempo ou algum mecanismo biológico (SEBER; WILD, 2003; PAINE et al., 2012). A flexibilidade desta curva para se ajustar aos dados vem de seus parâmetros, que possuem interpretação biológica (SEBER; WILD, 2003). Os modelos de Brody, Gompertz, logístico e von Bertalanffy (Tabela 2) são amplamente utilizados para descrever fenômenos biológicos (SEBER; WILD, 2003; UEDA et al., 2010; PAINE et al., 2012; ARCHONTOULIS; MIGUEZ, 2015).

Tabela 2 - Função, velocidade e aceleração dos modelos logístico, Brody, Gompertz e von Bertalanffy.

Modelo	Função	Velocidade	Aceleração
logístico	$Y = \frac{\beta_1}{1+C}$	$\frac{\partial Y}{\partial x} = \frac{\beta_1 \beta_3 C}{(1+C)^2}$	$\frac{\partial^2 Y}{\partial x^2} = \frac{\beta_1 \beta_3^2 C(C-1)}{(1+C)^3}$
Brody	$Y = \beta_1(1-D)$	$\frac{\partial Y}{\partial x} = \beta_1 \beta_3 D$	$\frac{\partial^2 Y}{\partial x^2} = -\beta_1 \beta_3^2 D$
Gompertz	$Y = \beta_1 \exp(-C)$	$\frac{\partial Y}{\partial x} = \beta_1 \beta_3 C \exp(-C)$	$\frac{\partial^2 Y}{\partial x^2} = -\beta_1 \beta_3^2 \exp(-C)(C-1)$
von Bertalanffy	$Y = \beta_1(1-D)^3$	$\frac{\partial Y}{\partial x} = 3\beta_1 \beta_3 D(1-D)^2$	$\frac{\partial^2 Y}{\partial x^2} = -3\beta_1 \beta_3^2 D(3D^2 - 4D + 1)$

* $C = \exp(\beta_2 - \beta_3 x)$; $D = \exp(-\beta_2 - \beta_3 x)$

Nos modelos Brody, Gompertz, logístico e von Bertalanffy o parâmetro β_1 corresponde a assíntota, que é o valor de estabilização da variável dependente; β_2 é um parâmetro de escala relacionado com a velocidade do crescimento inicial; e β_3 representa a taxa de crescimento (SEBER; WILD, 2003). Esses parâmetros possuem interpretação biológica que podem ajudar a descrever o comportamento produtivo das culturas olerícolas ao longo do tempo.

Todos os modelos acima são assintóticos. Porém diferem quanto ao comportamento da taxa de crescimento. No modelo de Brody a taxa de crescimento decai constantemente, dando ao modelo um aspecto não sigmoide. Já nos modelos de Gompertz, logístico e von Bertalanffy, a taxa de crescimento começa lenta, atinge um pico e volta a cair, dando um

aspecto sigmoide a curva (SEBER; WILD, 2003). A produção acumulada em olerícolas possui um comportamento sigmoide, o que permite utilizar os modelos de crescimento para descrever a produção de olerícolas (Figura 3A), sendo que vários trabalhos mostraram que esses modelos se ajustam aos dados de produção de olerícolas de múltiplas colheitas (LÚCIO et al., 2015; LÚCIO et al., 2016a; LÚCIO et al., 2016b).

2.5.1 Detalhando a taxa de crescimento dos modelos

Diferentemente do modelo de Brody, que tem uma taxa de crescimento que decresce constantemente, nos modelos Gompertz, logístico e von Bertalanffy a taxa de crescimento tem um comportamento distinto ao longo do tempo (SEBER; WILD, 2003). Esse comportamento pode ser determinado através da sua derivada de primeira ordem em relação a variável independente (Tabela 2). Plotando a derivada de primeira ordem em relação ao tempo, percebe-se que a taxa de crescimento é sempre positiva e tende a crescer até atingir um pico. Após atingir esse pico, a taxa de crescimento começa a decrescer (Figura 2B). O ponto em que a taxa de crescimento atinge o seu máximo é conhecido como ponto de inflexão (*PI*) (SEBER; WILD, 2003). Após atingir esse ponto, a velocidade decresce (Figura 3B) e a produção tende a se estabilizar (Figura 2A).

Como a taxa de crescimento representa a velocidade com que a produção aumenta ao longo do tempo é possível mensurar a sua aceleração através da derivada parcial de segunda ordem (MISCHAN et al., 2011; MISCHAN et al., 2015). A aceleração do crescimento determina o incremento na taxa de crescimento (aumento da velocidade = aceleração) ao longo do tempo. Diferentemente da taxa de crescimento ela pode ser negativa ou positiva. A aceleração de crescimento é positiva até atingir o ponto de inflexão e negativa após ele. Assintoticamente, a aceleração tende a zero (Figura 3C), representando ausência de incremento na taxa de crescimento e, conseqüentemente, a estabilização do modelo (Figura 3A). A derivada segunda possui um ponto de máxima, conhecido como ponto de aceleração máximo (PAM), e de mínimo, conhecido como ponto de desaceleração máxima (PDM). A fase de estabilização da aceleração determina um momento em que os incrementos na produção são insignificantes, conhecido como o ponto de desaceleração assintótica (PDA) (MISCHAN et al., 2011; PASSOS et al., 2012; MISCHAN et al., 2015).

As coordenadas do PAM, PI, PDM e PDA são obtidos igualando-se a zero as derivadas de segunda, terceira e quarta ordem, respectivamente (MISCHAN et al., 2011; PINHO et al., 2014; MISCHAN; PINHO, 2014) (Tabela 3). Esses pontos delimitam fases

chaves da produção de olerícolas, e são importantes para determinar a precocidade produtiva da cultura (conforme demonstrado mais adiante).

Tabela 3 - Coordenadas [x;y] dos pontos de aceleração máxima (PAM), inflexão (PI), desaceleração máxima (PDM) e desaceleração assintótica (PDA) obtidos matematicamente a partir dos modelos logístico, Brody, Gompertz e von Bertalanffy (MISCHAN; PINHO, 2014).

Modelo ¹	PAM	PI	PDM	PDA
logístico	$\left[\frac{\beta_2 - 1.317}{\beta_3}, \frac{\beta_1}{4.732} \right]$	$\left[\frac{\beta_2}{\beta_3}, \frac{\beta_1}{2} \right]$	$\left[\frac{\beta_2 + 1.317}{\beta_3}, \frac{\beta_1}{1.267} \right]$	$\left[\frac{\beta_2 + 2.292}{\beta_3}, \frac{\beta_1}{1.101} \right]$
Brody	-	-	-	-
Gompertz	$\left[\frac{\beta_2 - 0.962}{\beta_3}, \frac{\beta_1}{13.698} \right]$	$\left[\frac{\beta_2}{\beta_3}, \frac{\beta_1}{2.725} \right]$	$\left[\frac{\beta_2 + 0.962}{\beta_3}, \frac{\beta_1}{1.466} \right]$	$\left[\frac{\beta_2 + 1.797}{\beta_3}, \frac{\beta_1}{1.181} \right]$
von Bertalanffy	$\left[\frac{-\beta_2 + 0.303}{\beta_3}, \frac{\beta_1}{55.555} \right]$	$\left[\frac{-\beta_2 + 1.098}{\beta_3}, \frac{\beta_1}{3.378} \right]$	$\left[\frac{-\beta_2 + 1.894}{\beta_3}, \frac{\beta_1}{1.631} \right]$	$\left[\frac{-\beta_2 + 2.645}{\beta_3}, \frac{\beta_1}{1.247} \right]$

¹Baseado nas parametrizações apresentadas na Tabela 2. Fonte: MISCHAN; PINHO, 2014.

As coordenadas dos pontos críticos apresentados na tabela acima correspondem a parametrizações específicas dos modelos (ver Tabela 2). Porém, se o modelo for sigmoide, os pontos críticos também podem ser obtidos observando os pontos de máxima e mínima nas derivadas de primeira e segunda ordem em relação a x obtidas pela função *deriv()* e *deriv3()* do software R. Essa é uma forma prática de obter os valores que anulam as derivadas em qualquer modelo sigmoide.

2.5.2 Comparações de parâmetros

É comum que o pesquisador conduza experimentos com diversos fatores/fontes de variação (genótipos, ambientes, tratamentos), o que possibilita verificar a igualdade dos parâmetros entre eles. A forma mais comum de comparar parâmetros é pela abordagem modelo completo vs modelo reduzido, utilizando variáveis dummy's (FOX, 2002). Considere um experimento com três genótipos de uma espécie olerícola qualquer. Para testar a hipótese nula $H_0 : \beta_{11} = \beta_{12} = \beta_{13}$ dois modelos são propostos:

$$\begin{aligned}\omega: Y_{ij} &= \sum_{j=1}^3 \left[\frac{\beta_{1j}}{1 + e^{(\beta_{2j} - \beta_{3j}t_i)}} \right] + \varepsilon_{ij} \\ \Omega: Y_{ij} &= \sum_{j=1}^3 \left[\frac{\beta_1}{1 + e^{(\beta_{2j} - \beta_{3j}t_i)}} \right] + \varepsilon_{ij}\end{aligned}\quad (13)$$

Em (13), o índice subscrito $i = 1, 2, \dots, n$ corresponde às observações e $j = 1, 2$ e 3 refere-se aos genótipos. No modelo completo (ω), cada um dos parâmetros é estimado para cada um dos genótipos, totalizando nove parâmetros. No modelo reduzido, o parâmetro estimado $\hat{\beta}_1$ é comum aos três genótipos, e o modelo passa a ter apenas sete parâmetros.

A “perda” de parâmetros no modelo reduzido leva a um aumento na soma de quadrados do erro (SQ_{Erro}) e a um aumento nos graus de liberdade do erro (GL_{Erro}). Para que o modelo com menor número de parâmetros seja selecionado, o aumento da SQ_{Erro} deve ser compensada pelo aumento no GL_{Erro} , mantendo o quadrado médio do resíduo aproximadamente igual. Se houver essa compensação, conclui-se que poder de explicação de ambos os modelos é o mesmo e, portanto, o parâmetro β_j não difere entre os genótipos. Uma forma de verificar isso estatisticamente é através do teste F (SEBER; WILD, 2003; RITZ; STREIBIG, 2008; ARCHONTOULIS; MIGUEZ, 2015):

$$F_{calc} = \frac{[SQ_{Erro}(\Omega) - SQ_{Erro}(\omega)] / [GL_{Erro}(\Omega) - GL_{Erro}(\omega)]}{QM_{Erro}(\omega)} \quad (14)$$

Em (14), $SQ_{Erro}(\omega)$ e $SQ_{Erro}(\Omega)$ são as somas de quadrados dos resíduos nos modelos completo e reduzido, respectivamente; $GL_{Erro}(\omega)$ e $GL_{Erro}(\Omega)$ são os graus de liberdade do resíduo do modelo completo e reduzido, respectivamente; e $QM_{Erro}(\omega)$ é o quadrado médio do resíduo do modelo completo. O valor de F_{calc} é comparado com o quantil da distribuição F com p e $n-p-1$ graus de liberdade e nível α de significância [$F_{(\alpha, p, n-p-1)}$], onde α corresponde ao erro tipo I, p ao número de parâmetros do modelo e n ao número de observações. Se o valor de F_{calc} for maior que o quantil $F_{(\alpha, p, n-p-1)}$, rejeita-se H_0 (com α probabilidade de erro) e conclui-se que os parâmetros diferem entre os genótipos. A interpretação também pode ser realizada através do p -valor, sendo que valores menores que 0,05 indicam rejeição da hipótese H_0 com uma probabilidade de erro igual a 0,05.

Outra forma de testar a hipótese H_0 é pelo teste χ^2 , dado por (SEBER; WILD, 2003; RITZ; STREIBIG, 2008; ARCHONTOULIS; MIGUEZ, 2015):

$$\begin{aligned}\chi_{calc}^2 &= 2 \log \left(\frac{L_{\omega}}{L_{\Omega}} \right) \\ &= 2 (\log(L_{\omega}) - \log(L_{\Omega}))\end{aligned}\tag{15}$$

Em (15), $L(\omega)$ e $L(\Omega)$ são os valores máximos da função de máxima verossimilhança do modelo completo e reduzido, respectivamente. O valor de χ_{calc}^2 é comparado o quantil da distribuição χ^2 com ν graus de liberdade e nível α de significância $\chi_{(\alpha,\nu)}^2$, onde ν é a diferença no número de parâmetros entre o modelo completo (ω) e reduzido (Ω). Valores tabelados χ_{calc}^2 maiores que o quantil $\chi_{(\alpha,\nu)}^2$ indica a rejeição da hipótese H_0 com α probabilidade de erro. A interpretação também pode ser realizada através do *p-valor*, sendo que valores menores que 0,05 indicam rejeição da hipótese H_0 com uma probabilidade de erro igual a 0,05.

Para tamanhos de amostra grandes, as estatísticas F e χ^2 são equivalentes. Porém, quando a amostra é pequena, é aconselhável utilizar o teste F . Em modelos não lineares os testes F e χ^2 são aproximados, pois não possuem distribuição exata. Entre os dois testes é recomendável utilizar o teste F para comprar parâmetros, pois ele mantém o erro tipo I em um nível menor mesmo em amostras pequenas (REGAZZI; SILVA, 2004). Seber e Wild (2003) também mencionam que o teste F é mais poderoso.

Esses testes mencionados são amplamente utilizados para testar a hipótese em modelos aninhados. Mas a comparação também pode ser realizada com base no intervalo de confiança dos parâmetros. Também é possível estimar um modelo para cada repetição e utiliza-los como variáveis em um teste de Tukey. Ressalta-se que para realizar todas essas inferências é necessário o cumprimento dos pressupostos dos resíduos e que o modelo tenha uma boa aproximação linear (CARVALHO et al., 2010).

2.5.3 Utilização dos modelos de crescimento: uma abordagem com olerícolas

O ajuste adequado dos modelos de crescimento, levando em consideração os aspectos estatísticos já apresentados, possibilita ao pesquisador fazer inferências sobre o comportamento produtivo das culturas que vão muito além da produção final. Isso aumenta

significativamente os resultados que podem ser extraídos de um banco de dados e aumenta consideravelmente o conhecimento e as conclusões sobre a produção de olerícolas ao longo do seu ciclo produtivo.

O aumento das inferências só é possível graças a interpretação biológica das estimativas dos parâmetros desses modelos. Assim destaca-se a importância de não apenas selecionar os modelos pela sua qualidade de ajustamento, mas também levar em consideração as propriedades dos estimadores (aproximadamente não viesados). Logo, a qualidade das respostas obtidas está menos relacionada com a qualidade de ajuste do modelo e mais com o quanto as estimativas dos parâmetros são não viesadas. E esse é um aspecto fundamental que pesquisadores de culturas olerícolas deverão levar em consideração em suas pesquisas.

2.5.3.1 Interpretando os parâmetros e os pontos críticos

O modelo logístico parece ser o mais adequado para descrever o comportamento produtivo de olerícolas de múltiplas colheitas, por isso ele será utilizado como exemplo. Os incrementos nas estimativas do parâmetro β_1 causa mudanças nos valores da ordenada do modelo e das suas derivadas, alterando os valores da produção, taxa de crescimento e aceleração de crescimento. Porém os valores da abcissa, que determinam o momento em que os pontos críticos ocorrem, não são alterados (Figura 9). Mudanças nas estimativas do parâmetro β_2 causa um deslocamento ao longo da abcissa, impactando diretamente no momento em que os pontos críticos ocorrem. Porém, ele não altera os valores da produção, taxa de crescimento e aceleração de crescimento (Figura 10). Por fim, incrementos nas estimativas do β_3 causa um aumento na inclinação da curva de crescimento, e leva a um deslocamento tanto na ordenada como na abcissa das suas derivadas (Figura 11).

Percebe-se, também, que os pontos críticos delimitam fases importantes da produção das culturas olerícolas. O PI determina o momento em que a taxa de produção de frutos é máxima, ou seja, as plantas estão produzindo frutos plenamente. O PAM e o PDM delimitam a produção exponencial de frutos. Esta fase geralmente é curta, porém responsável por quase 60% da produção total de frutos (MISCHAN; PINHO, 2014). Antes do PAM e depois de PDM essa produção é muito lenta. Por fim, o PDA o momento em que a aceleração de crescimento tende a estabilidade (a taxa de crescimento tende a zero e, conseqüentemente, a produção se estabiliza). A partir desse ponto os incrementos na produção tendem a ser insignificante.

A partir do significado e do efeito dos que os parâmetros causam nos modelos, pode-se utilizá-los para determinar as seguintes características:

a) Produção total: o valor da estimativa do parâmetro β_1 será utilizado para determinar a produção final das culturas (Figura 9);

b) Precocidade produtiva: o valor da estimativa do parâmetro β_2 está relacionado com um aumento da produção lento ou mais rápido. Valores menores das estimativas indicam uma produção rápida, estando relacionado com um grau de maturação mais avançado das plantas no início das colheitas (Figura 10A). O ponto de inflexão (Figura 10B) e o período em que a produção é exponencial (Figura 10C) ocorrem mais precocemente. O parâmetro β_3 também está relacionado com a precocidade produtiva. Valores maiores do parâmetro demonstram produção mais precoce (Figura 11);

c) Período de concentração da produção: além de estar relacionado com a precocidade, o parâmetro β_3 também está relacionado com a concentração das colheitas. Valores estimados mais elevados resultam em a) inclinações mais acentuadas da curva de crescimento (Figura 11A), b) valores maiores de Y_{PI} e que ocorrem mais precocemente (menor X_{PI}) (Figura 11B) e c) colheitas mais precoces e com produção concentrado em um curto período de tempo (Figura 11C).

Figura 9 - Alteração (indicada pela seta) no modelo logístico (A), na derivada de primeira (B) e segunda ordem (C) com o incremento dos parâmetros β_1 .

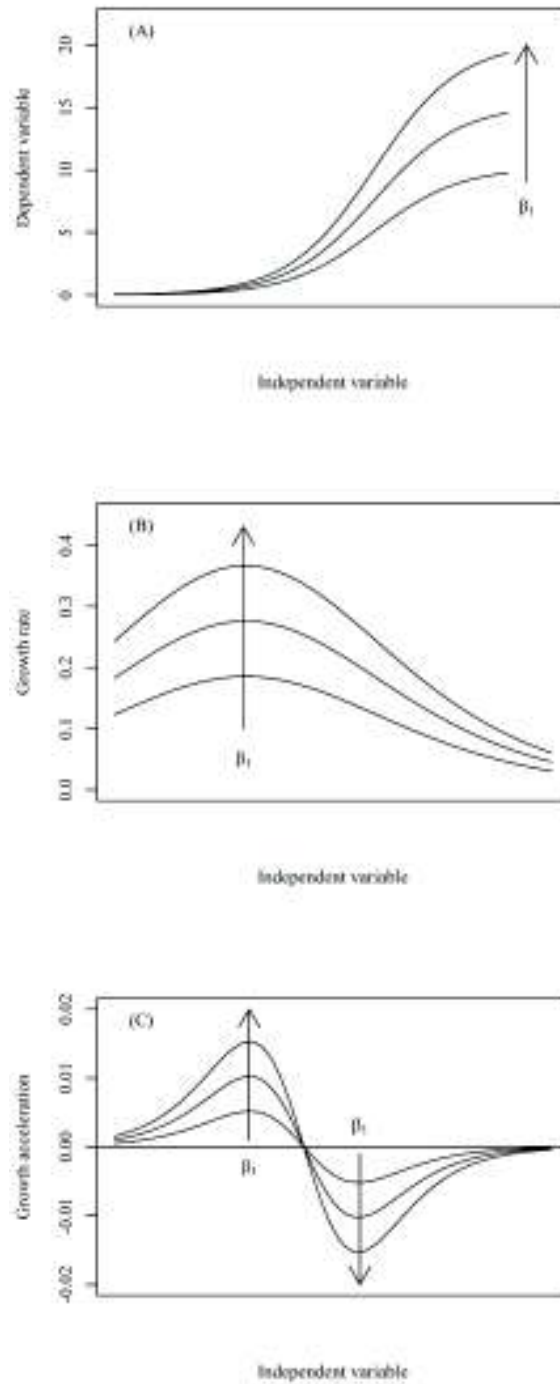


Figura 10 - Alteração (indicada pela seta) no modelo logístico (A), na derivada de primeira (B) e segunda ordem (C) com o incremento dos parâmetros β_2 .

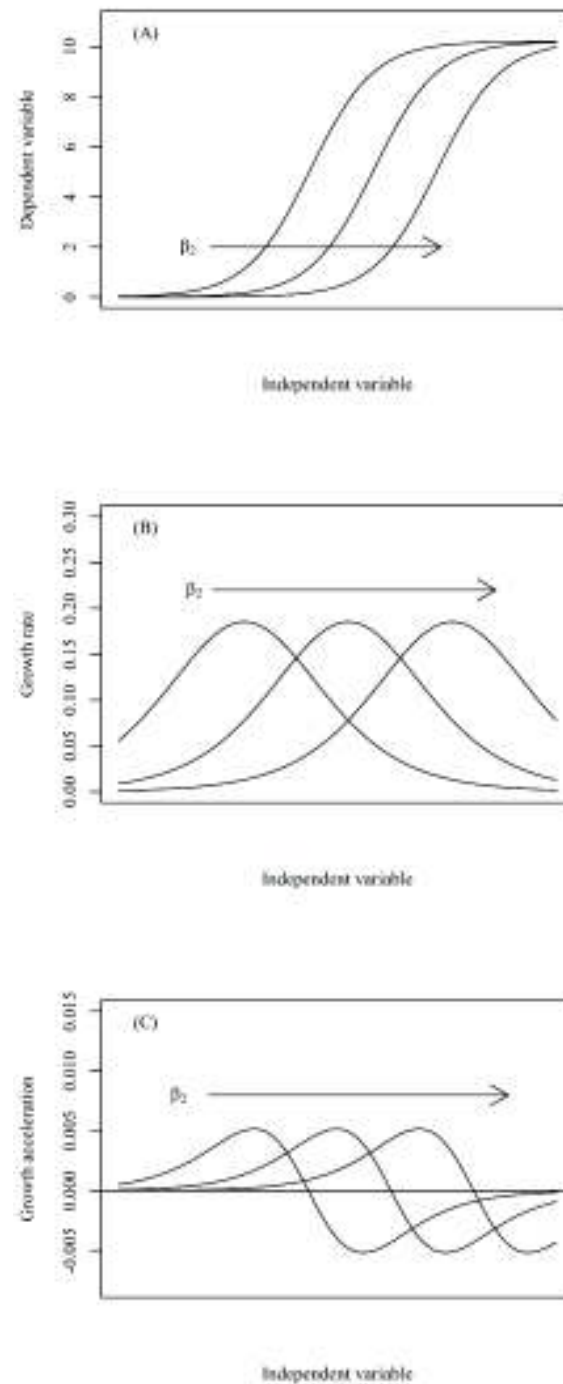
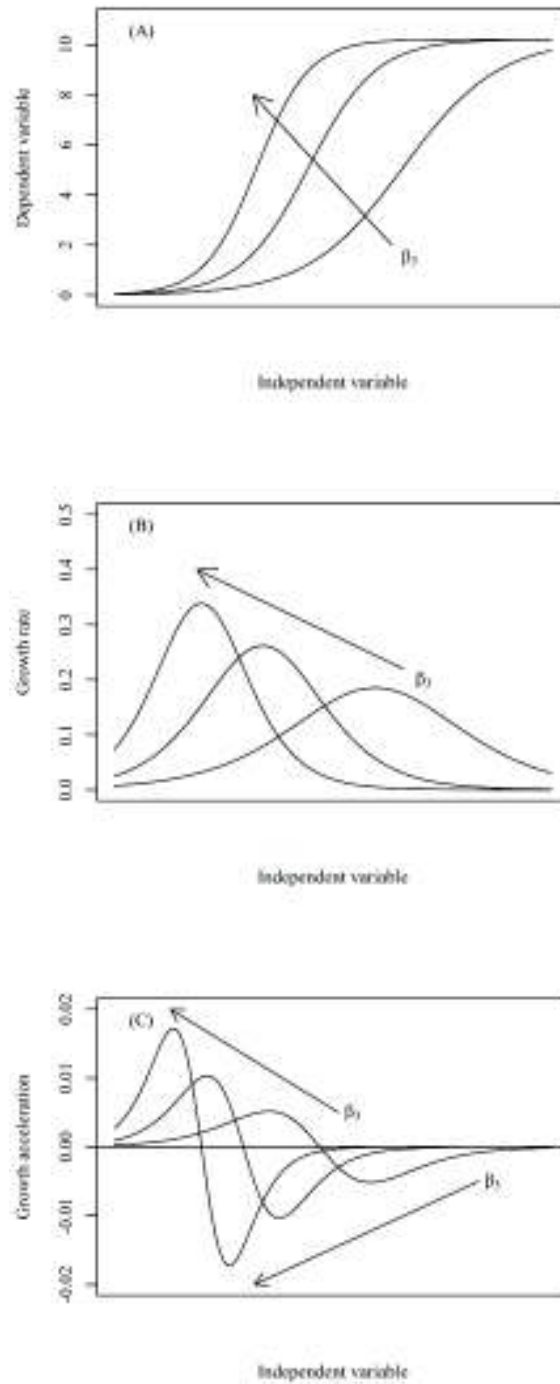


Figura 11 - Alteração (indicada pela seta) no modelo logístico (A), na derivada de primeira (B) e segunda ordem (C) com o incremento dos parâmetros β_3 .



1 **3 ARTIGO I - Describing tomato plant production using growth models**
2 **(Artigo nas normas da revista Scientia Horticulturae)**

3 Bruno Giacomini Sari^{a,*}, Alessandro Dal'Col Lúcio^a, Cinthya Souza Santana^a, Taciana Villela
4 Savian^b

5
6 ^a Federal University of Santa Maria, Department of Plant Science, Av. Roraima 1000, 97105-
7 900, Santa Maria, RS, Brazil

8 ^b University of São Paulo, College of Agriculture, Department of Math and Sciences, Av.
9 Pádua Dias 11, 13418-900, Piracicaba, SP, Brazil.

10 * Corresponding author.

11 E-mail address: brunosari@hotmail.com

12
13
14
15
16
17
18
19
20
21
22
23 *Abbreviations:* IP, inflection point; DAT, days after transplant of the seedlings; X_{PI} , abscissa
24 value for the inflection point; Y_{PI} , ordinate value of the inflection point; OLS, minimum
25 squares method.

26 Highlights

- 27 • The growth models describe the fruit production of tomato.
- 28 • Logistic model is appropriate to model the fruit production of tomato genotype.
- 29 • Growth models increase the inference about productive behavior.
- 30 • We define production precocity by parameters and critical points interpretation.
- 31 • The production precocity was defined with statistical rigor and not subjectively.

32

33 Abstract

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

51 **Keywords:** *Solanum lycopersicum*; Non-linear models; Production growth; Productive
52 precocity; Vegetable crops.

53

54 **Introduction**

55 Some vegetable crops provide more than one harvest over the crop cycle and are
56 known as multiple harvest crops. One of the consequences of this characteristic is that under
57 experimental conditions, production must be measured separately for each harvest, as the
58 variability between the plants will be different for each of these (Lúcio et al., 2015; Lúcio and
59 Sari, 2017). This occurs because the maturation of plant is not uniform. This factor, together
60 with overdispersion and the high number of zero values in the database (Lúcio and Sari,
61 2017), causes recurring problems in experimental analysis, such as the failure to meet the
62 assumptions of the statistical model in analysis of variance.

63 Using harvests (repeated measurements) as a fixed factor can lead to violation of the
64 independence of residuals in analysis of variance (ANOVA), since harvests cannot be
65 randomized in the experiment (Fernandez, 2007). In addition, the presence of overdispersion
66 and/or zero-inflated data also leads to violation of assumptions such as normality and
67 homoscedasticity of residuals, making it impossible to evaluate the data by conventional
68 ANOVA (Fernandez, 2007, Lúcio and Sari, 2017). One way to circumvent these problems is
69 to group plants and harvests so that there is a reduction in heterogeneity and the assumptions
70 of ANOVA are met. Therefore, it is common for analysis of data from multi-harvest crops to
71 be carried out based on total production only.

72 Despite the problems cited above, regression models can be used as a statistical data
73 analysis tool for multiple harvests in vegetable crops (Lúcio et al., 2015, Lúcio et al., 2016a,
74 Lúcio et al., 2016b). As the yield values measured for plants in each harvest accumulate, crop
75 production begins to display sigmoidal behavior (Seber and Wild, 2003, Paine et al., 2012),

76 which starts slowly and grows exponentially, before dropping and stabilizing. This type of
77 behavior is typical in non-linear regression models, known as “growth models”. They contain
78 biologically interpreted parameters that quantify the speed of the growth of an organism and
79 determine when this growth tends to stabilize (Seber and Wild, 2003, Mischan et al., 2011).

80 The use of non-linear models to describe growth is common in agricultural science. In
81 vegetable crops with multiple harvests, Wubs et al. (2012) used growth models to describe the
82 growth of pepper fruits. However, this statistical method is still uncommon in vegetable crops
83 and is mostly limited to adjusting growth models to accumulated production data. For
84 example, Lúcio et al. (2015), Lúcio et al. (2016a) and Lúcio et al. (2016b) used the logistic
85 and von Bertalanffy models to describe fruit production in zucchini, pepper, cherry tomato
86 and green beans, and obtained increasing quality adjustments. However, these studies only fit
87 the models to the data, and do not use parameters and critical points as variables to describe
88 vegetable crop yield.

89 Similarly, the non-linear models can be used to describe the photosynthesis response
90 to different environment, to modeling the temperature dependence of a plant process and to
91 describe a rate of phenological development (Archontoulis and Miguez, 2015). Olivoto et al.
92 (2018) and Elli et al. (2018) used the Schumacher’s model to describe the accumulated dry
93 matter in escarole (*Cichorium endivia* L. var. *latifolia*) grown. As with growth models, these
94 studies only fit the models to the data.

95 Since growth models variables involve biological interpretation, these estimates can
96 also be used to compare behavior of production among treatments. According to Seber and
97 Wild (2003), Ritz and Streibig (2008) and Archontoulis and Miguez (2015), dichotomous
98 variables can be used to verify whether a set of models has the same parameter values via the
99 F test, which compares parameters between models. Puiatti et al. (2013) used model identity
100 in this way to verify the similarity of the development of bulbs in garlic genotypes.

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

109 Thus, besides describe the productive behavior of tomato plant, the growth models
110 (through its parameters and critical points) can be used to determine subjective characteristics,
111 such as productive precocity. In this way, the productive precocity can be defined exactly in
112 function of the productive behavior, and a mathematical solution can be given to a
113 characteristic that is determined subjectively (in most of the case).

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

120 The sigmoidal behavior of production over time allows growth models to be adjusted,
121 and can be used as an important statistical tool, increasing the number of inferences regarding
122 productive behavior, based on an interpretation of model parameters, and overcoming the
123 limitations of the ANOVA. The aim of this study is to describe the productive behavior of
124 tomato genotypes (*Solanum lycopersicum*) using adjusted growth models, and determine the
125 productive precocity through the parameters and critical points interpretation.

126

127 **Materials and Methods**

128

129 *Site description and experimental design*

130 Data was taken from two field experiments performed at the Department of Plant
131 Science at the Federal University of Santa Maria (UFSM), Rio Grande do Sul state, Brazil
132 (29°43'S, 53°43'O, altitude 95 m) during 2015/2016 and 2016/2017. The Cordillera, Ellen
133 and Santa Clara hybrids were used for the 2015/2016 experiment and the Cordillera and
134 Gaucho hybrids were used for the 2016/2017 experiment. The seedlings were transplanted on
135 October 17, 2015, in the first experiment, and on September 10, 2016, in the second
136 experiment. The spacing of rows between the seedlings were 1.00 and 0.30 m for the
137 2015/2016 experiment and 1.00 and 0.50 m for the 2016/2017 experiment. The plants were
138 staked using polypropylene cord and a single stake in both experiments. Fertilizing was based
139 on chemical analysis of the soil and followed recommendations by the Official Network of
140 Soil and Plant Tissue Analysis Laboratories for the States of Rio Grande do Sul and Santa
141 Catarina (ROLAS, 2004). Pest and disease management followed the technical
142 recommendations for the crop. Climatic data was collected by a weather station situated
143 200 m from the experiment location.

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

149

150 *Adjustment and selection of models*

151 The measured values for weight and quantity of fruit per plant for each harvest were
 152 accumulated for each genotype, after which the following growth models were adjusted: a)
 153 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:
 154 $Y_i = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 t_i)}} + \varepsilon_i$; 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
 155 of the accumulated variables (number and weight of fruit) over time (dependent variable); t_i is
 156 the time interval between transplant and the harvest (independent variable); β_1 is the
 157 parameter representing the asymptotic value; β_2 reflects the distance between the initial and
 158 the asymptotic value; and β_3 is associated with the rate of growth.

159 Parameters were estimated using the ordinary least-squares method, using a Gauss-
 160 Newton approach and solving ordinary least square at each iteration. Following this, the
 161 Shapiro-Wilk, Breusch-Pagan and Durbin-Watson tests were applied to verify the normality,
 162 homogeneity and the independence of the residuals respectively. The quality of the fit of each
 163 adjusted model for each genotype was determined using the coefficient of determination (R^2)
 164 and through verification of the nonlinearity of the model, using the Bates and Watts curvature
 165 method (Bates and Watts, 2007). The model that presented the highest R^2 value (close to 1)
 166 and low nonlinearity values (intrinsic < 0.3 and parametric < 1.0) was selected.

167

168 *Determination of the inflection point*

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

174

175

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

176

177

178

179

Where X_{PI} is the abscissa of the inflection point, $t_{(0.05;GL_{Error})}$ is the value of the of 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:

180

181

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

182

183

184

185

186

187

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 $Cov(\hat{\beta}_2, \hat{\beta}_3)$ respectively, these being the estimates for the variance (and covariance) for these parameters.

188

189

190

191

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 dummy (D_j):

192

$$D_j = \begin{cases} 1, & \text{if the observation pertence to genotype } k; \\ 0, & \text{in any other case.} \end{cases}$$

193

194

195

Where j is the genotype evaluated. The purpose of this test was to compare a complete model (ω), presenting different parameters for each genotype, with a reduced model (Ω),

196 presenting similar parameters between genotypes. Using the logistic model as an example, the
 197 complete model would be as follows:

198

$$199 \quad Y_{ij} = \sum_{j=1}^k \left[\frac{\beta_{1j}}{1 + e^{(\beta_{2j} - \beta_{3j}t_i)}} \right] + \varepsilon_{ij} \quad (3)$$

200

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

204

$$205 \quad Y_{ij} = \sum_{j=1}^k \left[\frac{\beta_1}{1 + e^{(\beta_{2j} - \beta_{3j}t_i)}} \right] + \varepsilon_{ij} \quad (4)$$

206

207 Not that in (4), the parameter β_1 is the same for all k genotypes. The null hypothesis
 208 would be $H_0: \beta_{11} = \beta_{12} = \dots = \beta_{1k}$. Using the methodology described above, it was verified whether
 209 the null hypotheses were rejected or not for both variables, considering a significance level of
 210 5%:

211

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

212

213
 214 A comparison of the complete and reduced models was performed using the F test
 215 (Ritz and Streibig, 2008). The value of F_{calc} was calculated as follows:

216

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

218

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

223 The parameters effect on the selected model was represented graphically, from
 224 simulated data, to facilitate the interpretation of the results. For this, graphs were constructed
 225 varying only one of the model parameters, keeping the others constant. The productive
 226 behavior of the genotypes over time was represented in graph form in order to facilitate
 227 interpretation of the results. All analyses were performed using R software (R Core Team,
 228 2014).

229

230 **Results and discussion**

231

232 *Model selection*

233 All the models fit the weight and number of fruit per plant data for the experiment
 234 conducted in 2015/2016. For the experiment conducted in 2016/2017, the Brody model could
 235 not be adjusted to the two genotypes evaluated (Tables 1 and 2). The assumptions of
 236 normality, homoscedasticity and independence of the residuals were met, indicating that the
 237 estimation of the parameters using the ordinary least-squares method was adequate, as
 238 recommended by Ritz and Streibig (2008).

239

240

241

242 **Table 1**

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

Genotypes	----- N° of fruits -----					----- Weight of fruit -----				
	SW	BP	DW	R^2	c^θ	SW	BP	DW	R^2	c^θ
----- Brody -----										
Cordillera	0.95	0.09	0.71	0.96	444.47 ¹	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

247 ¹ Values obtained using the equation $c^\theta \times \sqrt{F_{(5\%,3,3)}}$.

248

249 **Table 2**

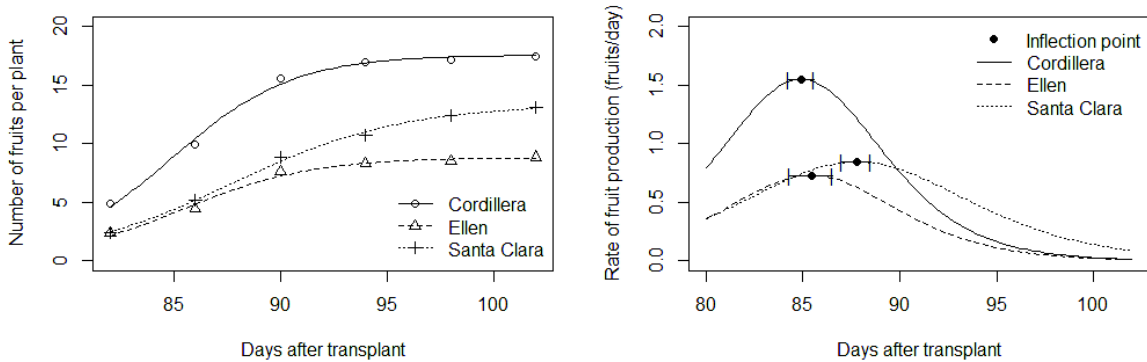
250 P-values for the Shapiro-Wilk (SW), Breusch-Pagan (BP) and Durbin-Watson (DW) tests
 251 applied to the residuals, the coefficient of determination (R^2) and the parametric nonlinearity
 252 (c^θ) of the Brody, Gompertz, logistic and von Bertalanffy models in three tomato genotypes,
 253 for the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

Genotype	----- N° of fruits -----					----- Weight of fruit -----				
	SW	BP	DW	R^2	c^θ	SW	BP	DW	R^2	c^θ
----- Brody -----										
Cordillera	- ¹	-	-	-	-	-	-	-	-	-
Gaicho	-	-	-	-	-	-	-	-	-	-
----- Gompertz -----										
Cordillera	0.34	0.10	0.13	0.99	18.21 ²	0.89	0.12	0.15	0.99	17.57
Gaicho	0.45	0.67	0.52	0.99	19.80	0.74	0.51	0.98	0.99	29.45
----- Logistic -----										
Cordillera	0.45	0.46	0.23	0.99	0.81	0.79	0.37	0.28	0.99	0.60
Gaicho	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
Gaicho	0.94	0.20	0.57	0.99	15.31	0.31	0.21	0.30	0.99	11.94

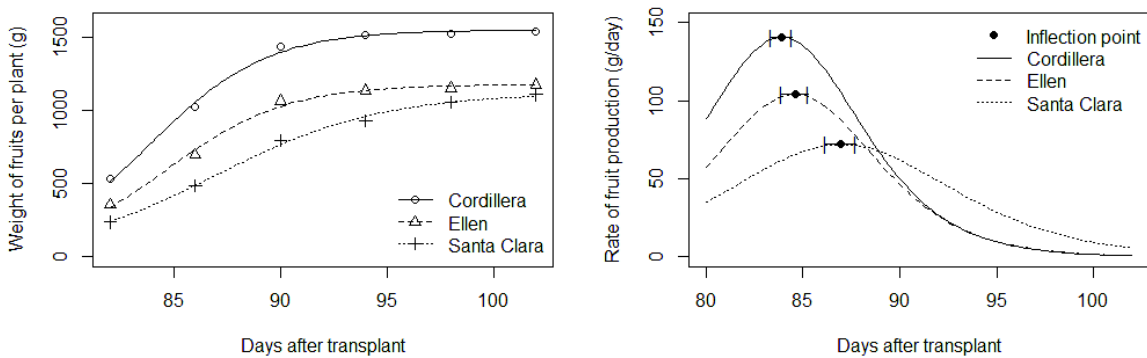
254 ¹ The Brody method did not fit the data. ² Values obtained using the equation $c^\theta \times \sqrt{F_{(5\%,3,6)}}$.

255

256 There was a difference in productive behavior between the genotypes for the two
 257 years. The first point to be highlighted is the time that harvesting began, which occurred later
 258 for the first experiment than for the second. In 2015/2016, the degree of ripening for the
 259 genotypes was higher when harvesting began. This reflects the absence of slower initial
 260 growth and the possibility of the Brody model fitting the data from the first experiment. This
 261 model represents data for which the growth rate tends to decrease from the start, and this
 262 behavior seems to be more plausible for the first rather than the second experiment (Figures 1
 263 and 2).

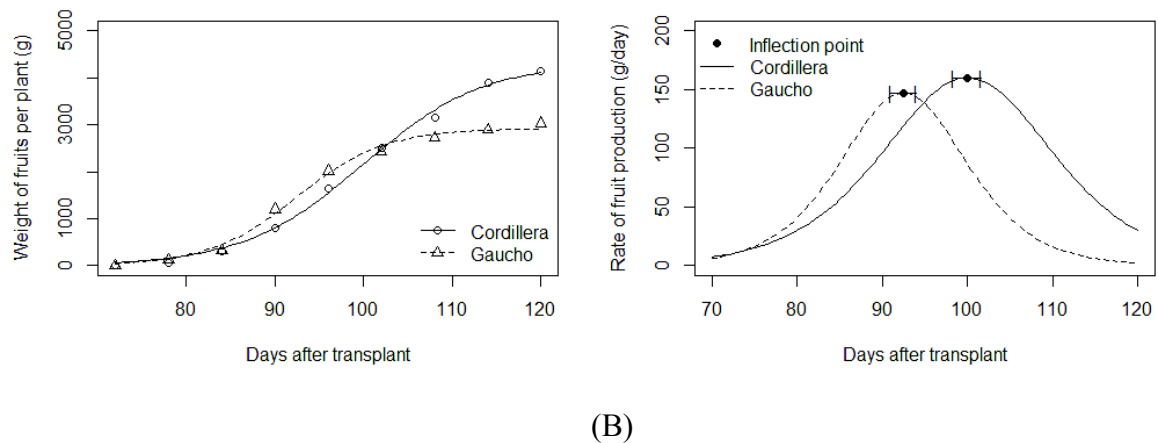
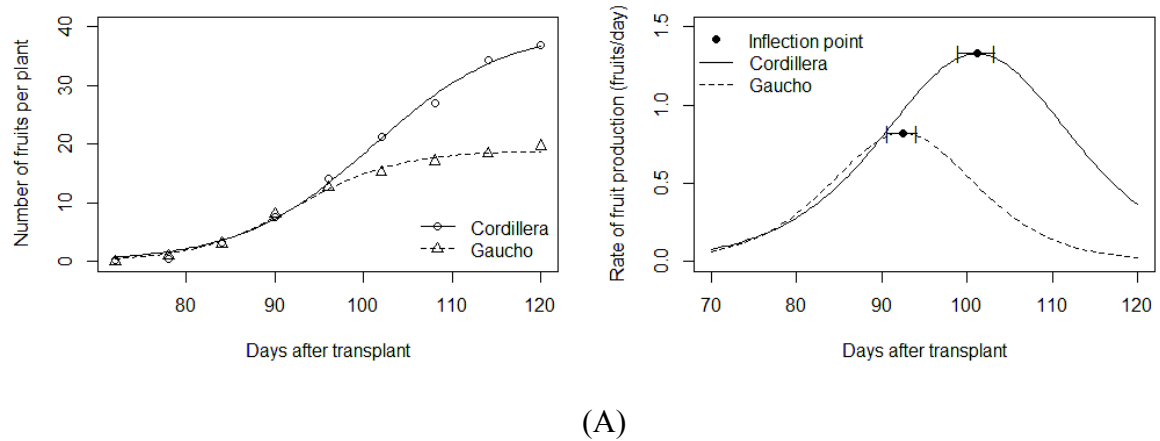


(A)



(B)

264 **Fig. 1.**
 265 Observed value (points), adjusted logistic model (lines), rate of fruit production and 95%
 266 confidence interval for the inflection point of three tomato genotypes for the variables number
 267 (A) and weight of fruit per plant (B) in 2015/2016.
 268



269 **Fig. 2.**
 270 Observed values (points), adjusted logistic model (lines), rate of fruit production and 95%
 271 confidence interval for the inflection point of three tomato genotypes for the variables number
 272 (A) and weight of fruit per plant (B) in 2016/2017.
 273

274 Despite all of the models having met the assumptions of normality, homogeneity and
 275 independence of the residuals, and all of them demonstrating high quality adjustment (high
 276 R^2), the logistic model presented the lowest parametric nonlinearity (Tables 1 and 2).
 277 Therefore, independent of the difference in behavior between the experiments mentioned
 278 above, the logistic model is the most adequate in describing the production data on the
 279 genotypes evaluated. As the estimates for the non-linear model parameters are obtained via a
 280 linear approach using the least-squares method, the behavior of the non-linear model should
 281 be close to that of the linear model. In this case, the estimators take (approximately) on the

282 properties of linear model parameters, are more (approximately) unbiased and estimates for
283 biologically inferred estimative of the parameters are more reliable (Seber and Wild, 2003;
284 Bates and Watts, 2007; Ritz and Streibig, 2008). The value of the curvature that determines if
285 a non-linear model has parametric low linearity or not is controversial, but values below 1 are
286 acceptable (Seber and Wild, 2003).

287 All of the adjusted models presented a high degree of adjustment (high R^2) and lower
288 intrinsic nonlinearity (< 0.3), however the logistic method presented the lowest curvature due
289 to the effect of the parameters (low parametric nonlinearity measures) (Seber and Wild, 2003;
290 Bates and Watts, 2007), and was selected for this reason (Tables 1 and 2). Since the
291 parameters are to be used to describe and compare the productive behaviors of the genotypes,
292 the high adjustment of a non-linear model (determined by high R^2) should also include
293 estimates of (approximately) unbiased parameters (in line with the real observations).
294 Therefore selecting growth models purely on the basis of their high adjustment can lead to
295 erroneous conclusions if the parameter values are not open to biological interpretation.

296

297 *Interpretation of parameters*

298 The estimated values for the asymptote β_1 of the model can be considered to be the
299 estimates for total production for the genotypes evaluated. For the 2015/2016 crop, the
300 Cordillera genotype produced an average of 17 fruits per plant, with a total weight of
301 1549.72 g, differing statistically from the other genotypes. The average weight of fruit per
302 plant for the Ellen and Santa Clara genotypes of around 1100 g did not differ statistically.
303 However, the average number of fruits per plant was higher for the Santa Clara genotype
304 (approximately 13 fruits per plant) compared to the Ellen genotype (approximately 8 fruits per
305 plant), demonstrating that the average weight of the fruit of the latter genotype is higher than
306 the former. For the 2016/2017 crop, the Cordillera genotype produced an average of 39 fruits

307 per plant, with a total weight of 4320.10 g, differing statistically from the Gaucho genotype,
 308 which produced an average of 19 fruits per plant and had a total weight of 2921.10 g (Tables
 309 3 and 4).

310

311 **Table 3**

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

Genotype	N° of fruits					Weight of fruit				
	β_1	β_2	β_3	X_{PI}^1	Y_{PI}^2	β_1	β_2	β_3	X_{PI}	Y_{PI}
----- 2015/2016 -----										
Cordillera	17.58*	29.98*	0.35*	84.91	8.79	1549.72*	30.45*	0.36*	83.90	774.86
Ellen	8.84*	28.22*	0.33*	85.42	4.42	1178.19*	29.88*	0.35*	84.58	589.09
Santa Clara	13.35*	22.24*	0.25*	87.76	6.67	1118.52*	22.39*	0.26*	86.95	559.26
----- 2016/2017 -----										
Cordillera	39.68*	13.54*	0.13*	101.12	19.84	4320.10*	14.82*	0.15*	99.96	2160.05
Gaucho	18.94*	16.04*	0.17*	92.42	9.47	2921.10*	18.65*	0.20*	92.44	1460.55

315 *Non-zero parameters according to the Student t test, $P < 0.05$. ¹ Abscissa of the inflection
 316 point;

317 ² Ordinate of the inflection point.

318

319

320

321

322

323

324

325

326

327

328

329

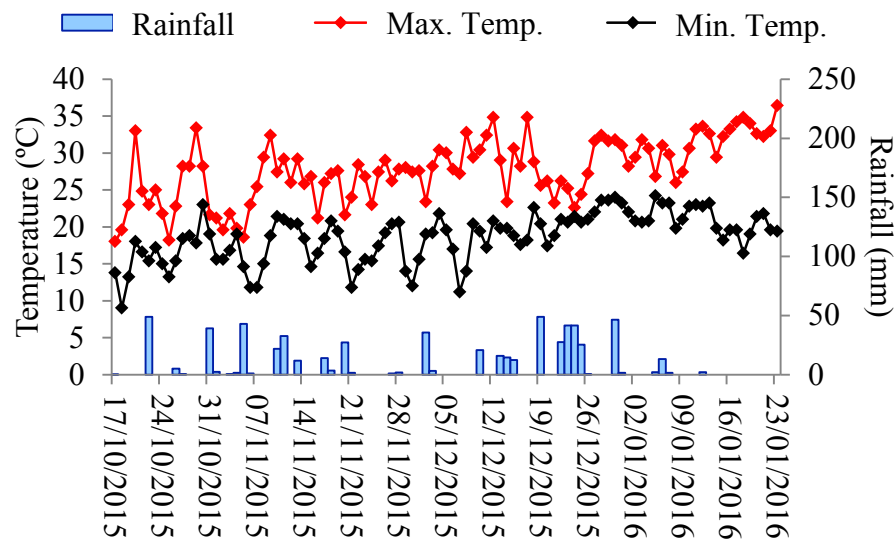
330 **Table 4**
 331 P-values for the F test to evaluate the identity of the logistic model between tomato
 332 genotypes, for the variables of number and weight of fruit (in grams) per plant, for 2015/2016
 333 and 2016/2017.

Hypotheses	Year	Number of fruits	Weight of fruit
		<i>p</i> -values	
----- β_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
----- β_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
----- β_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

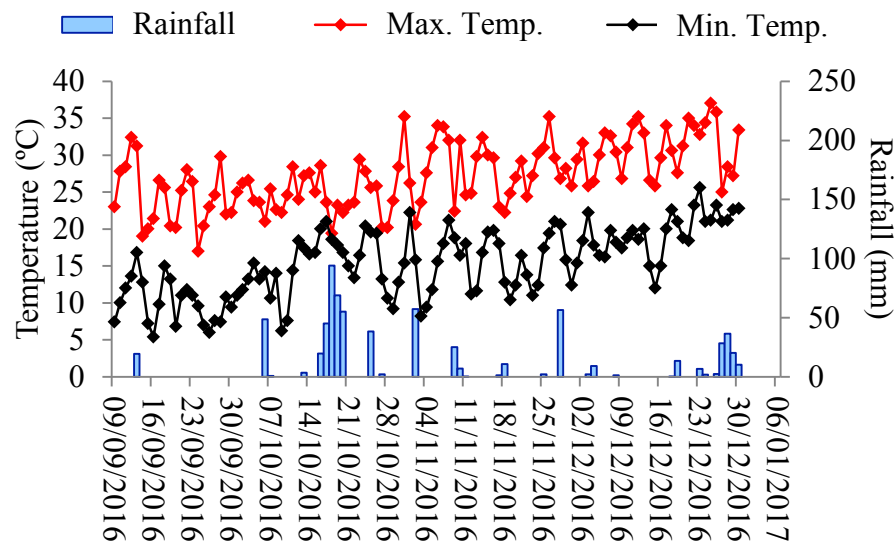
334 * $\beta_{.1}$ = Parameters for the Cordillera genotype; $\beta_{.2}$ = parameters for the Ellen genotype; $\beta_{.3}$ =
 335 parameters for the Santa Clara genotype; $\beta_{.4}$ = parameters for the Gaucho genotype.
 336

337 The lower production in 2015/2016 occurred due to excessive rainfall during the
 338 fructification stage of the crop, while in 2016/2017, the rains occurred mainly during the
 339 vegetation period (Figure 3). This together with the fact that the plant population was higher
 340 for the 2015/2016 experiment compared to the 2016/2017 experiment explains this large
 341 difference in terms of weight and number of fruits per plant for the two years. Note that
 342 average yield per hectare was equal to or higher than that observed in the state of Rio Grande
 343 do Sul, Brazil (IBGE, 2017).

344



(A)



(B)

345 **Fig. 3.**
 346 Maximum temperature (Max. Temp.), minimum temperature (Min. Temp.) and daily rainfall
 347 for 2015/16 (A) and 2016/17 (B).
 348

349 The parameter β_1 are related to the total production of the genotypes, as already
 350 reported. The change in the estimates of β_1 does not shift the curve in relation to the abscissa
 351 axis (Figure 4A) and also does not affect the moment of PI occurrence (Figure 5A). The lower
 352 estimate of β_2 and higher estimate of β_3 are related to the early production of the genotypes.
 353 The relationship of the parameters β_2 and β_3 with the early production is given by the shift (in

354 relation to abscissa) and greater slope of the growth curve (Figures 4B and 4C). It is also
355 possible to verify the early production by the moment of PI occurrence (X_{PI} values) (Figures
356 5B and 5C). The X_{PI} is the time at which of fruit production rate is maximal. As the logistic
357 model is symmetrical (Seber and Wild, 2003), X_{PI} is also the time when 50% of genotype
358 production has already been harvested.

359 When the estimate of β_2 is smaller (keeping the other parameters constant) the growth
360 curve shifts to the left, showing that the production increase tends to begin earlier (Figure 4B).
361 Similarly, the PI occurs earlier (lower X_{PI}) when the estimate of β_2 is smaller (Figure 5B). The
362 same behavior is observed when the estimative of the β_3 is increase. The growth curve shifts
363 to the left due to the higher slope (higher fruit production rate), and that this indicate the
364 increase in production is higher and occurs earlier (Figure 4C). Similarly, the PI occurs earlier
365 (lower X_{PI}) when the estimate of β_3 are higher (Figure 5C). The parameter β_3 associate the
366 early production with the higher production rate, whereas the parameter β_2 does not (Figure
367 5B and 5C).

368

369

370

371

372

373

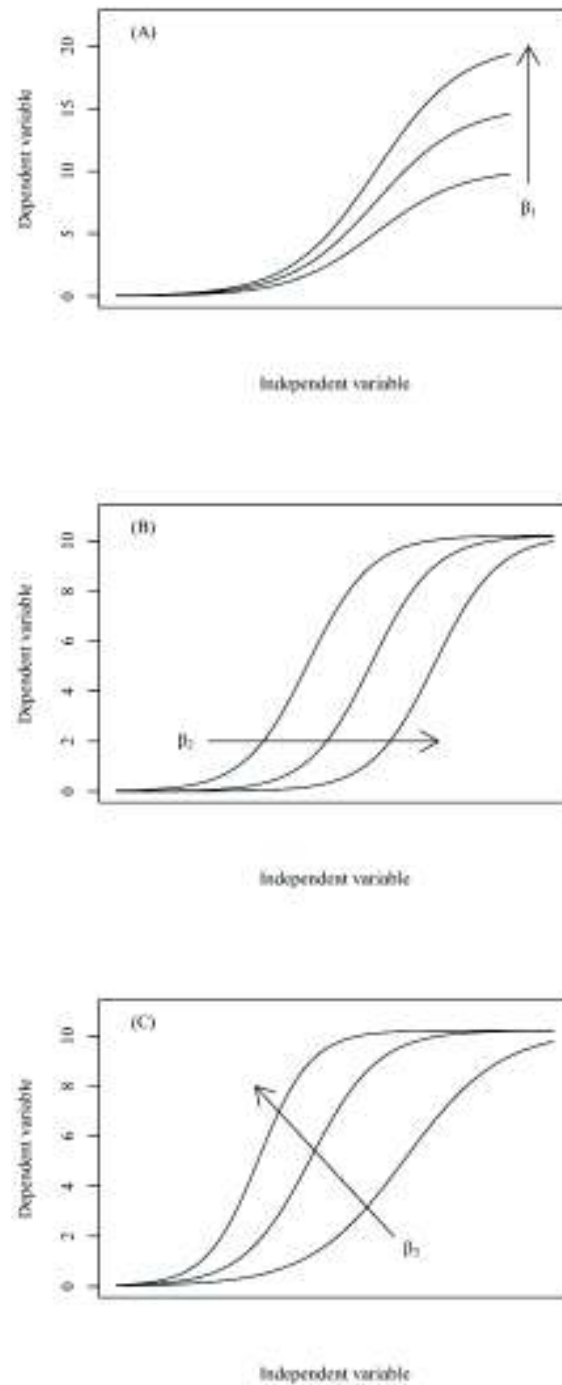
374

375

376

377

378



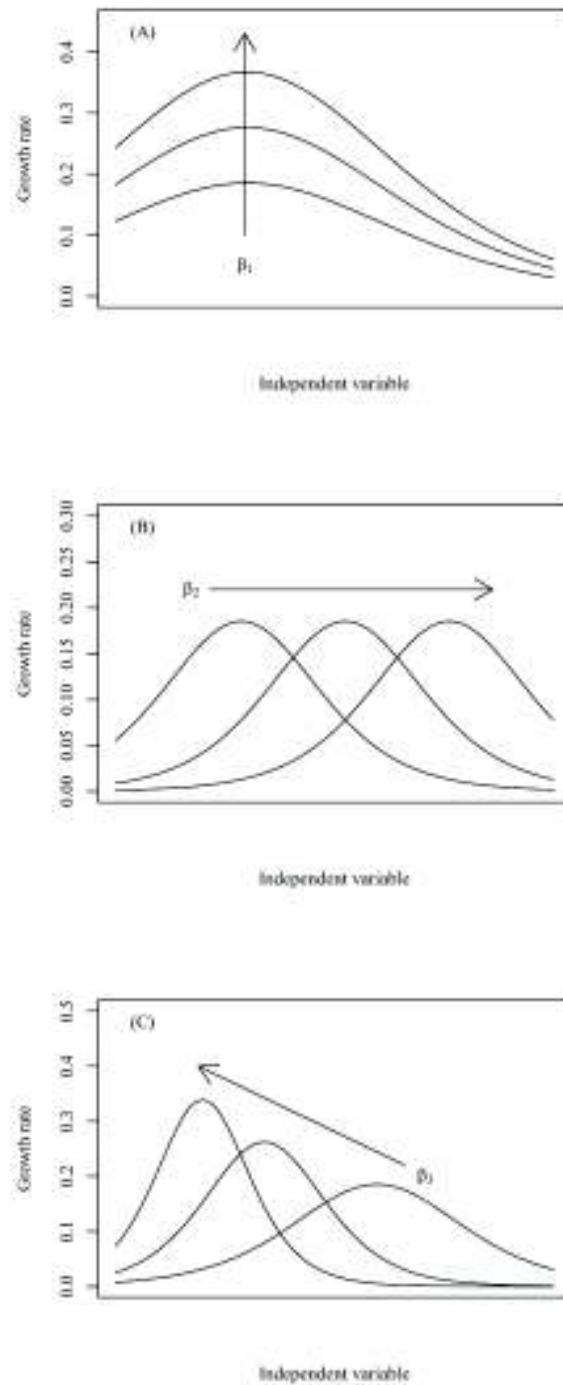
379

380 **Fig. 4.**381 Effect of increases in parameters β_1 (A), β_2 (B) and β_3 (C) on the logistic model:382
$$Y_i = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 t_i)}} + \varepsilon_i$$
. Arrows indicate changes in the curves as the parameter values increase,

383 all other constants remaining the same.

384

385



386

387 **Fig. 5.**388 Effect of increases in parameters β_1 (A), β_2 (B) and β_3 (C) on the growth rate in the logistic389 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

390 increase, all other constants remaining the same.

391

392 The parameters β_2 and β_3 should be analyzed together. As reported above, lower
393 estimate of β_2 and higher estimate of β_3 indicate earlier production. When both estimates of β_2
394 e β_3 ere high, as in case of this study (Table 3), it was concluded that the fruit production is
395 early. Therefore, the higher fruit production rate (β_3) is the component of the model that has
396 greater influence on earlier production in this case (Figures 1 and 2).

397 For the 2015/2016 crop, the comparison of Cordillera and Santa Clara genotypes
398 shows well the relationship between β_2 and β_3 parameters. For the Cordillera genotype, the
399 estimates of β_2 and β_3 were statistically higher in relation to the Santa Clara genotype (Table 3
400 and 4). Although the higher estimates values of β_2 indicate otherwise, the Cordillera genotype
401 produced fruits earlier (earlier occurrence of PI) (Figure 1). For the 2015/2016 crop, the same
402 is observed. Although the higher estimates values of β_2 indicate otherwise (Table 3), the
403 Gaucho genotype produced fruits earlier (Figure 2). Therefore, the productive precocity of
404 these genotypes is related to the higher estimate value of β_3 .

405 The point of the cycle at which the rate of production is at its maximum is represented
406 by the abscissa value for the inflection point (X_{PI}). For the 2015/2016 crop, X_{PI} did not differ
407 statistically between the Cordillera and Ellen genotypes for either variable. For these
408 genotypes, the inflection point for both the weight and the number of fruits variables was
409 reached at approximately 85 DAT. For the Santa Clara genotype, the inflection point was
410 reached at a later stage for both variables, at around 87 DAT (Table 3 and Figure 1). For the
411 2016/2017 crop, there was a statistical difference between X_{PI} values. The Gaucho genotype
412 reached the inflection point for both variables at approximately 93 DAT, which was earlier
413 than the Cordillera genotype, which reached the inflection point at around 100 DAT (Table 3
414 and Figure 2).

415 From simulated data, the relationship between lower estimates of β_2 and higher
416 estimation of β_3 with the productive precocity of the genotypes was verified. The shift of the

417 growth curve to the left (Figure 4B), the higher slope (Figure 4C) and the earlier occurrence
418 of the PI (Figure 5B and 5C) were used in this interpretation. It is clear that the productive
419 precocity, represented by the earliest occurrence of the PI (lower X_{PI}), is related to the higher
420 estimates of β_3 in this study.

421 In addition to determining total production through the parameter β_1 , the growth
422 models increase the inferences that can be made through the interpretation of the other
423 parameters (mainly β_3 in this study) and the determination PI coordinates on the curve.
424 Similarly, comparisons between nested models and estimated confidence intervals for the
425 critical points allow treatments to be compared (Ritz and Streibig, 2008; Mischan et al., 2011;
426 Pinho et al., 2014). This way it was possible to make inferences on the productive precocity
427 of the tomato genotypes evaluated. In 2015/2016, the Cordillera and Ellen genotypes were
428 more premature than the Santa Clara genotype. For the 2016/2017 crop, the Gaucho genotype
429 tended to be more premature than the Cordillera genotype (Tables 3 and 4 and Figures 2 and
430 3).

431

432 *Use of growth models as a statistical analysis tool*

433 For the statistical analysis of the experiments, the use of the logistic model determined
434 total production for the genotypes and enabled comparisons to be made on other aspects of the
435 productive behavior of the tomato plants. However, growth models can be used in other ways,
436 and not just as a statistical analysis tool. For example, they can be used to get a better
437 understanding of the productive behavior of vegetables over harvest periods and describe this
438 through growth curves (and their derivatives), can be used to help researchers or agricultural
439 extension agents to recommend planting periods for genotypes to avoid overloading
440 workforces during determined points of the harvest period, or to plan production in terms of

441 commercialization (Figures 2 and 3). Wider studies, including tests into genotype and
442 transplant period, can contribute towards technical recommendations for crops.

443 Growth models have proven to be efficient statistical analysis tools. They have been
444 used to make inferences regarding genotype productive behavior based on yield data that can
445 be easily collected in the field and reported according to the number of days after harvesting.
446 Furthermore, model parameters and critical points (especially β_3 and X_{PI}) can be used to
447 identify the genotypes with the highest yield, predict the time of highest yield and therefore
448 the time at which the yield is greatest. In other words, productive behavior can be quantified
449 by the model's parameter estimate values and its inflection point, which can be used as
450 explanatory variables to compare genotypes. This information is important, but is largely
451 ignored due to measurement problems. We have shown that, using the correct statistical
452 approach, these problems can be overcome.

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

462 To obtain the nonlinear parameters estimative by OLS, the models must be
463 approximately linear (Seber and Wild, 2003; Bates and Watts, 2007). Thus, the low
464 nonlinearity measures indicate that the parameters estimates are (approximately) unbiased. As
465 the parameter estimates have biological interpretation, and determine the PI and its confidence

466 interval, the selection of the models must be performed by the low nonlinearity measures (that
467 indicate approximately unbiased estimates). In this way, the model's selection must be
468 performed by these measures, and not only by the numerical statistics that measure a
469 goodness of fit (e.g., R^2) or selection criteria measures (e.g., AIC, BIC). Using only the
470 goodness of fit measures, the models selection can be have wrong.

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

476

477 *Defining productive precocity*

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

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

491

492 **Conclusion**

493 The growth models made possible describe the behavior of the tomato plant
494 production. The selection of logistic model was performed by the lower nonlinearity
495 measures. The β_1 estimates represents the total production of the genotypes. By the β_3
496 estimate and of the inflection point, it was possible determine the genotypes production
497 precocity. In this study, the Cordillera and Ellen genotypes were more premature in
498 2015/2016 crop, and the Gaucho genotype was more premature in the 2016/2017 crop.

499

500 **Acknowledgement**

501 We thank the Brazilian National Council for Scientific and Technological
502 Development (CNPq) and the Coordination for the Improvement of Higher Level Personnel
503 (CAPES), for granting scholarships to the authors.

504

505 **References**

- 506 Archontoulis, S.V., Miguez, F.E., 2015. Nonlinear Regression Models and Applications in
507 Agricultural Research. *Agron. J.*, 107, 786–798
508 <https://doi.org/10.2134/agronj2012.0506>.
- 509 Bates, D.M., Watts, D.G., 2007. *Nonlinear Regression Analysis and its Applications*. John
510 Wiley and Sons Inc., New York, pp. 392.
- 511 Elli, E.P., Olivoto, T., Schmidt, D., Caron, B.O., Souza, V.C., 2018. Precision of growth
512 estimates and sufficient sample size: can solar radiation level change these factors?
513 *Agron. J.*, 110, 155–163. <https://doi.org/10.2134/agronj2017.05.0297>.
- 514 Fernandez, G.C.J. 2007. Design and analysis of commonly used comparative horticultural
515 experiments. *HortScience*, 42, 1052-1069.

- 516 IBGE – Instituto Brasileiro de Geografia e Estatística. Levantamento sistemático da produção
517 agrícola, março de 2017, 2017. Instituto Brasileiro de Geografia e Estatística, Rio de
518 Janeiro, pp. 83.
- 519 Jones, J.W., Dayan, E., Allen, L.H., Van Keulen, H., Challa, H., 1991. A dynamic tomato
520 growth and yield model (TOMGRO). *ASAE*, 34, 663-672.
521 <https://doi.org/10.13031/2013.31715>.
- 522 Lúcio, A.D., Nunes, L.F., Rego, F., 2015. Nonlinear models to describe production of fruit in
523 *Cucurbita pepo* and *Capiscum annuum*. *Sci. Hortic.*, 193, 286-293.
524 <https://doi.org/10.1016/j.scienta.2015.07.021>.
- 525 Lúcio, A.D., Sari, B.G., Rodrigues, M., Bevilaqua, L.M., Voss, H.M.G., Copetti, D., Faé, M.,
526 2016a. Modelos não-lineares para a estimativa da produção de tomate do tipo cereja.
527 *Cienc. Rural*, 46, 233-241. <https://doi.org/10.1590/0103-8478cr20150067>
- 528 Lúcio, A.D., Nunes, L.F., Rego, F., 2016b. Nonlinear regression and plot size to estimate
529 green beans production. *Hortic. bras.*, 34, 507-513. [http://dx.doi.org/10.1590/s0102-](http://dx.doi.org/10.1590/s0102-053620160409)
530 [053620160409](http://dx.doi.org/10.1590/s0102-053620160409)
- 531 Lúcio, A.D., Sari, B.G., 2017. Planning and implementing experiments and analyzing
532 experimental data in vegetable crops: problems and solutions. *Hortic. bras.*, 35, 316-
533 327. <http://dx.doi.org/10.1590/s0102-053620170302>
- 534 Mischan, M.M., Pinho, S.Z., Carvalho, L.R., 2011. Determination of a point sufficiently close
535 to the asymptote in nonlinear growth functions. *Sci. Agric.*, 68, 109-114.
536 <https://doi.org/10.1590/S0103-90162011000100016>.
- 537 Olivoto, T., Elli, E.P., Schmidt, D., Caron, B.O., Souza, V.C., 2018. Photosynthetic photon
538 flux density levels affect morphology and bromatology in *Cichorium endivia* L. var.
539 *latifolia* grown in a hydroponic system. *Sci. Hortic.*, 230, 178–185. [https://doi.org/](https://doi.org/10.1016/j.scienta.2017.11.031)
540 [10.1016/j.scienta.2017.11.031](https://doi.org/10.1016/j.scienta.2017.11.031).

- 541 Paine, C.E.T., Marthews, T.R., Vogt, D.R., Purves, D., Rees, M., Hector, A., Turnbull, L.A.,
542 2012. How to fit nonlinear plant growth models and calculate growth rates: an update
543 for ecologists. *Methods Ecol. Evol.*, 3, 245–256. [https://doi.org/ 10.1111/j.2041-](https://doi.org/10.1111/j.2041-210X.2011.00155.x)
544 [210X.2011.00155.x](https://doi.org/10.1111/j.2041-210X.2011.00155.x).
- 545 Pinho, S.Z., Carvalho L.R., Mischan, M.M., Passos, J.R.S., 2014. Critical points on growth
546 curves in autoregressive and mixed models. *Sci. Agric.*, 71, 30-37.
547 <https://doi.org/10.1590/S0103-90162014000100004>.
- 548 Puiatti, G.A., Cecon, P.R., Nascimento, M., Puiatti, M., Finger, F.L., Silva, A.R., Nascimento,
549 A.C.C., 2013. Análise de agrupamento em seleção de modelos de regressão não lineares
550 para descrever o acúmulo de matéria seca em plantas de alho. *Rev. Bras. Biom.*, 31,
551 337-351.
- 552 R Core Team. R: A language and environment for statistical computing, 2014. R Foundation
553 for Statistical Computing, Austria.
- 554 Ritz, C., Streibig, J.C., 2008. *Nonlinear Regression with R*. Springer, New York, pp. 142.
- 555 ROLAS - Rede Oficial de Análise de Solo e de Tecido Vegetal. Manual de adubação e
556 calagem para os estados do Rio Grande do Sul e Santa Catarina, 2004. Sociedade
557 Brasileira de Ciência do Solo, Porto Alegre, pp. 400.
- 558 Seber, G.A.F., Wild, C.J., 2003. *Nonlinear regression*. John Wiley and Sons Inc., New Jersey,
559 pp. 768.
- 560 Wubs, A.M., Ma, Y.T., Heuvelink, E.P., Hemerik, L., Marcelis, L.F.M., 2012. Model
561 Selection for Nondestructive Quantification of Fruit Growth in Pepper. *J. Amer. Soc.*
562 *Hort. Sci.*, 137, 71–79.

25 **Highlights**

- 26 • Nonlinear growth models were evaluated in tomato production data.
- 27 • Model derivatives allowed identifying production behavior
- 28 • Growth models allow for increase the type of inferences in tomato experiments.
- 29 • Growth models are an alternative tool to ANOVA models.
- 30 • The concentration of production was determined by the critical points.
- 31 • The precocity and concentration of production was determined statically and not
- 32 subjectively.

33

34 **Abstract**

35 The aim of this work was to use growth models as an alternative analysis of experiments with
36 tomatoes. The data were obtained from field experiments of tomatoes carried out in the
37 2015/2016 and 2016/2017 growing seasons. Six and nine harvests were carried out in
38 2015/2016 and 2016/2017, respectively. In each harvest the number and mass of fruits per
39 plant were assessed. For each variable, the Brody, Gompertz, Logistic and von Bertalanffy
40 growth models were fitted as a function of the accumulated values per plant in each harvest
41 and days after the transplant of the seedlings. The assumptions of normality, homogeneity and
42 independence of residuals were verified by the Shapiro-Wilk, Breusch-Pagan and Durbin-
43 Watson tests, respectively. The intrinsic and parametric nonlinearity was calculated according
44 to Bates and Watts methodology. Among the models tested, the one that presented a high
45 coefficient of determination and less nonlinearity were selected. The parameters of the
46 selected model were compared by the F test. From the second, third and fourth order
47 derivatives with respect to x (days after transplant), the maximum acceleration point,
48 inflection point, maximum deceleration point, and point of asymptotic deceleration were
49 estimated. All assumptions regarding residual distributions were met. Among the tested

50 models, the logistic model had the best goodness-of-fit for both variables. Therefore, it was
51 possible to test hypotheses to compare the parameters by the F test, increasing the inferential
52 value of the results. In addition, the critical points allowed assessing of the productive
53 precocity, crop concentration period and productive behavior of the genotypes during the
54 productive period. These results demonstrate that growth models may be an alternative to
55 analysis of variance in experiments with horticultural crops.

56 **Keywords:** *Solanum lycopersicum*; Nonlinear regression; Productive precocity; Productive
57 behavior; Logistic model.

58

59 **Introduction**

60 Multi harvest crops are characterized by having more than one harvest throughout
61 their productive cycle. These crops present a series of characteristics that must be considered
62 when analyzing the results of research. For example, the harvests are carried out in the same
63 plant more than once, being classified as repeated measures; these crops are also characterized
64 by uneven fruit maturation and high plant variability, commonly resulting in data with
65 overdispersion, zero-inflated data at the beginning and at the end of the harvests, and
66 heteroscedasticity between rows and between harvests (Lúcio et al., 2016a, 2015).

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

74 common for analysis of data from multi-harvest crops to be carried out considering total
75 production only.

76 By analyzing the total production only, several important portions of information
77 through the crop cycle are lost; consequently, the inferences are narrowed. However, aiming
78 at solving this problem, alternative statistical approaches to ANOVA can be utilized.
79 Nonlinear regressions are particularly appropriate, since when accumulating the values of
80 each harvests, it is observed that a sigmoidal curve can be fitted for the total production
81 (Lúcio et al., 2016b; Lúcio et al., 2015). Among nonlinear regressions, growth models are
82 widely used in research in the field of agricultural sciences (Bi et al., 2012; Carneiro et al.,
83 2014; Goldberg and Ravagnolo, 2015; Helidoniotis et al., 2011; Katsanevakis and Maravelias,
84 2008; Leduc and Goelz, 2009; Mehtätalo et al., 2015; Mendonça et al., 2014; Wubs et al.,
85 2012).

86 Nonlinear regression models also have assumptions to be satisfied. The nonlinearity of
87 the model and the homoscedasticity, normality and independence of the residuals are
88 necessary for the estimation of precise parameters and for the construction of confidence
89 intervals and hypotheses testing. The homoscedasticity and independence of the residuals can
90 be violated, since horticultural crops have high heterogeneity. However, these problems can
91 be overcome by estimating the parameters by the weighted least squares method, or by using
92 autoregressive errors, using the generalized least squares estimation method (Seber and Wild,
93 2003; Ritz and Streibig, 2008).

94 In addition to solving the statistical problem in data analysis of horticultural crops,
95 growth models can increase the inferential utility of the results, since these models have
96 parameters and critical points with a biological interpretation. The asymptote of the curve can
97 be used to determine the total production, and the growth parameter can be used to make
98 inferences about the speed (rate) of fruit production. Other points of the growth curve, such as

99 the inflection point and the points of acceleration and deceleration (Mischan et al., 2011;
100 Pinho et al., 2014; Seber and Wild, 2003) can help researchers to make inferences regarding
101 productive behavior over time, the precocity of production and the concentration of harvests.
102 According to Regazzi (2003), Seber and Wild (2003), Ritz and Streibig (2008) and
103 Archontoulis and Miguez (2015), the use of dichotomous variables (dummy variables) allows
104 for verifying whether or not a set of models has the same parameter value through the F test,
105 which allows for comparing the parameters of the models.

106 The benefits of using growth models in agronomic sciences are still little explored.
107 Their use in experiments with horticultural crops is particularly promising, as they can
108 increase the inferences of the results and circumvent statistical shortcomings of other
109 approaches that arise due to characteristics of the crop. In this regard, the present study aims
110 to verify if the growth models fit the tomato production data (multiple crop harvests), and if
111 they can be used as an alternative statistical tool to ANOVA.

112

113 **Material and methods**

114 The data were obtained from two tomato field trials carried out in Department of Plant
115 Science of the Federal University of Santa Maria (UFSM), RS, Brazil (29 ° 43'S, 53 ° 43'O,
116 95 m above sea level) in the 2015/2016 and 2016/2017 growing season (GS). The Cordillera,
117 Ellen, and Santa Clara hybrids were used in the 2015/2016 experiment and the Cordillera and
118 Gaucho hybrids were used in the 2016/2017 experiment. The transplanting of the seedlings
119 was carried out on October 17, 2015, in the first experiment, and on September 10, 2016 in
120 the second experiment. The spacing between rows and between the seedlings were 1.00 and
121 0.30 m in the 2015/2016 experiment, and 1.00 and 0.50 m in the 2016/2017 experiment,
122 respectively.

123 The training was carried out with plastic wires and the plants were conducted using
 124 wood single rods in both experiments. Fertilization was carried out based on the soil chemical
 125 analysis and according to the recommendation of the Official Network of Soil and Tissue
 126 Analysis Laboratories of the States of Rio Grande do Sul and Santa Catarina (ROLAS, 2004).
 127 Pest and disease management followed the technical recommendation of the crop. The
 128 climatic data were collected at an official meteorological station 200 m from the experiment
 129 site.

130 Harvestings were carried out at 82, 86, 90, 94, 98 and 102 days after transplant (DAT)
 131 in 2015/2016 and at 72, 78, 84, 90, 96, 102, 108, 114 and 120 DAT at 2016/2017. The fruits
 132 were harvested when they changed from green to reddish, and were later packed in bags and
 133 taken to the laboratory where they were counted and weighed on a digital scale (with
 134 milligram scale). The mean values of weight and number of fruits per plant, in each harvest,
 135 were accumulated within each genotype and the following growth models were fitted: (i)
 136 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:
 137 $Y_i = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 t_i)}} + \varepsilon_i$; and (iv) von Bertalanffy: $Y_i = \beta_1 (1 - \beta_2 e^{(\beta_3 t_i)})^3 + \varepsilon_i$, where: Y is the
 138 dependent variable (number and weight of fruits) accumulated through the DAT; t_i is the time
 139 from DAT to harvest; β_1 is the parameter that represent the asymptotic value; β_2 is the
 140 parameter that represent the distance between the initial value and the asymptote (in relation
 141 to the abscissa); β_3 is the parameter related to the growth rate.

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

148 model was one that presented the greater R^2 value (close to one) and less nonlinearity value
 149 (intrinsic<0.3 and parametric<1.0).

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

153

$$154 \quad D_j = \begin{cases} 1, & \text{if the observation belongs to } j \text{ genotypes;} \\ 0, & \text{if else.} \end{cases}$$

155

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

160

$$161 \quad Y_{ij} = \sum_{j=1}^k \left[\frac{\beta_{1j}}{1 + e^{(\beta_{2j} - \beta_{3j}t_i)}} \right] + \varepsilon_{ij}$$

162

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

166

$$167 \quad Y_{ij} = \sum_{j=1}^k \left[\frac{\beta_1}{1 + e^{(\beta_{2j} - \beta_{3j}t_i)}} \right] + \varepsilon_{ij}$$

168

169 It is observed that in this case, the parameter β_1 is common for all the k -genotypes.
 170 Thus, the formulated hypothesis in this example would be: $H_0: \beta_{11} = \beta_{12} = \dots = \beta_{1k}$, vs H_1 : not all

171 β_{1k} are equal. According to this example, for both variables, the following null hypothesis
 172 was tested assuming a 0.05 error probability.

173

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

175

176 The comparison between the complete and reduced models was performed by using
 177 the F test (Ritz and Streibig, 2008). The value of F was calculated by:

178

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

180

181 It was considered that F_{calc} follow a F distribution with $F_{[v, \text{DF}_{\text{Error}}(\omega)]}$ under H_0 , where v
 182 is the difference between DF_{Error} of the complete (Ω) and reduced (ω) models.

183 Finally, the coordinates of the critical points were obtained by using the partial
 184 derivatives of the selected model in relation to the independent variable (days after
 185 transplantation). The inflection point (IP) was obtained by equating the second order
 186 derivative to zero; the maximum acceleration (MAP) and deceleration (MDP) points were
 187 obtained by equating the third order derivative of the model to zero; and the asymptotic
 188 deceleration point (ADP) was obtained by equating the derivative of the fourth order to zero
 189 (Mischan et al., 2011). The curves of the model and the derivatives of the first and second
 190 order were represented graphically with the purpose of assisting in the interpretation of
 191 results. All analyses were performed in R software (R Core Team, 2014).

192

193 **Results**

194 All models, with the exception of the Brody model in the 2016/2017 GS, had good fit
 195 to the production data of the evaluated genotypes. Thus, the results for the Brody model in the
 196 2016/2017 GS are not presented. It was also observed that the assumptions of normality,
 197 homoscedasticity and independence of residuals were not violated. All the models showed
 198 high goodness-of-fit measures; however the logistic model presented an intrinsic nonlinearity
 199 less than 0.3 and parametric nonlinearity less than 1, and was therefore chosen (Tables 1 and
 200 2).

201

202 **Table 1**

203 P-values for Shapiro-Wilk (SW), Breusch-Pagan (BP) and Durbin-Watson (DW) tests,
 204 determination coefficient (R²), and nonparametric linearity ($c\theta$) of the Brody, Gompertz,
 205 Logistic and von Bertalanffy models in three tomato genotypes for the number and weight of
 206 fruits per plant in 2015/2016 growing season.

Genotypes	Number of fruits					Weight of fruits				
	SW	BP	DW	R ²	$c^{\theta a}$	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

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

208

209

210

211

212 **Table 2**
 213 P-values for Shapiro-Wilk (SW), Breusch-Pagan (BP) and Durbin-Watson (DW) tests,
 214 determination coefficient (R²), and nonparametric linearity (c⁰) of the Brody, Gompertz,
 215 Logistic and von Bertalanffy models in three tomato genotypes for the number and weight of
 216 fruits per plant in 2016/2017 growing season.

Genótipos	Number of fruits					Weight of fruits				
	SW	BP	DW	R ²	c ^{0a}	SW	BP	DW	R ²	c ⁰
Gompertz										
Cordillera	0.34	0.10	0.13	0.99	18.21	0.89	0.12	0.15	0.99	17.57
Gaúcho	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
Gaúcho	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
Gaúcho	0.94	0.20	0.57	0.99	15.31	0.31	0.21	0.30	0.99	11.94

217 ^avalues obtained by the expression $c^0 \times \sqrt{F_{(5\%,3,6)}}$.

218

219 The total production of the genotypes can be inferred from the asymptotic value. In the
 220 2015/2016 GS, the Cordillera genotype produced approximately 17 fruits per plant, totaling
 221 1549.72 g per plant. The Ellen genotype produced 8 fruits per plant, \cong 38% less than the Santa
 222 Clara genotype. The average weight per plant was the same, around 1100 g. In the 2016/2017
 223 GS, the Cordillera genotype produced an average 39 fruits per plant, with a total weight of
 224 4320.10 g, differing statistically from the Gaúcho genotype, which produced 19 fruits per
 225 plant with a total weight of 2921.10 g (Figures 1 and 2).

226

227

228

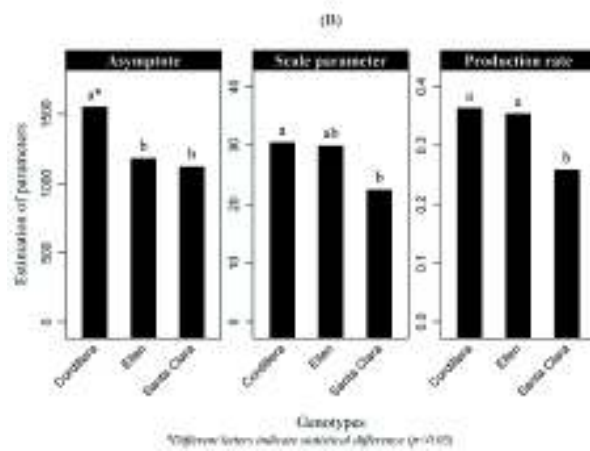
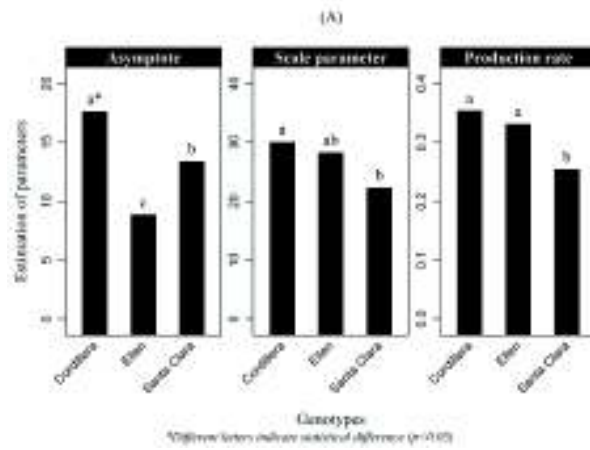
229

230

231

232

233



234

235

236

Fig. 1

237

Parameters of the logistic model for the number (A) and mass (B) of fruits per plant in three

238

tomato genotypes grown in 2015/2016.

239

240

241

242

243

244

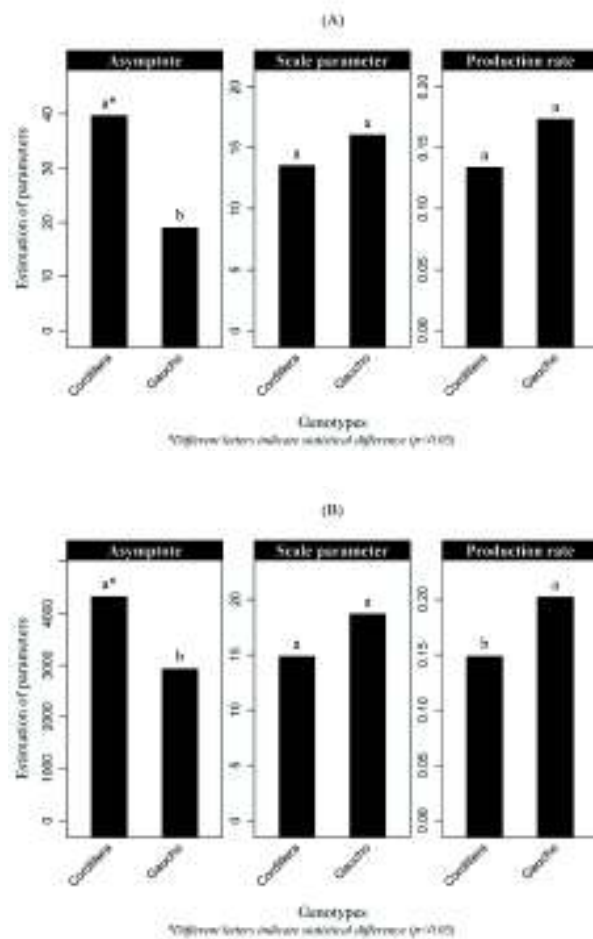
245

246

247

248

249



250

251

Fig. 2

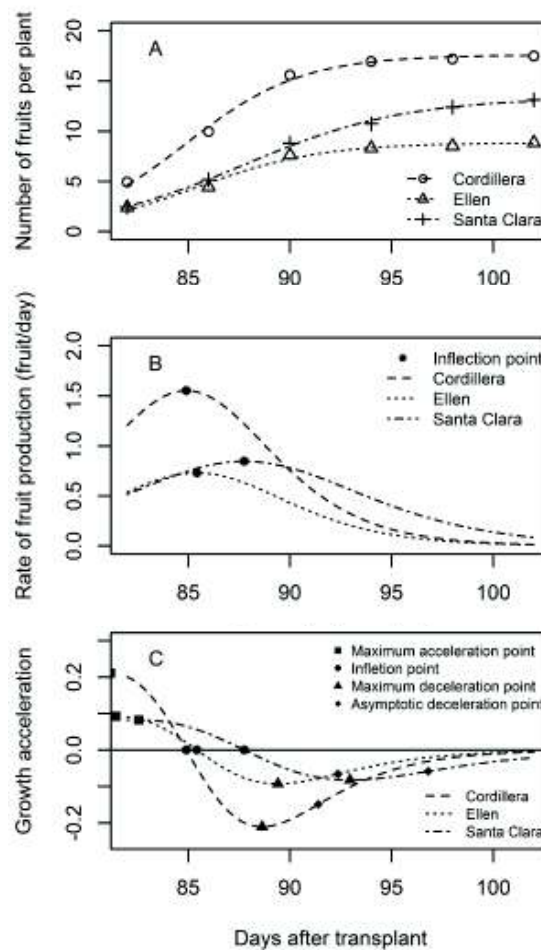
252 Parameters of the logistic model for the number (A) and mass (B) of fruits per plant in three
 253 tomato genotypes grown in 2016/2017.
 254
 255

256 The Cordillera genotype had an estimated value of β_2 (scale parameter) equal to the
 257 Ellen genotype and superior to the Santa Clara genotype, while the latter two did not differ
 258 (Figure 1). In the 2016/2017 GS, there was no statistical difference between the Cordillera
 259 and Gaúcho genotypes for both variables (Figure 2). Regarding the fruit production rate (β_3),
 260 significantly higher values in the Cordillera and Ellen were observed in relation to the Santa
 261 Clara genotype in 2015/2016 GS (Figure 1). In the 2016/2017 GS, the parameters differed
 262 statistically only for weight of fruits, and the rate of production was higher in the Gaúcho
 263 genotype (Figure 2).

264 The maximum acceleration point (MAP) indicates the moment at which the increase
 265 (acceleration) in the rate of fruit production (velocity) is maximal. The beginning of the

266 harvests occurred almost concomitantly with MAP in the 2015/2016 harvest. Therefore, the
 267 start was late with plants already producing large amounts of fruits in all genotypes,
 268 differently from what occurred in the 2016/2017 harvest, where the harvests started earlier
 269 and before reaching MAP (Figures 3 to 6).

270



271

272 **Fig. 3**

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

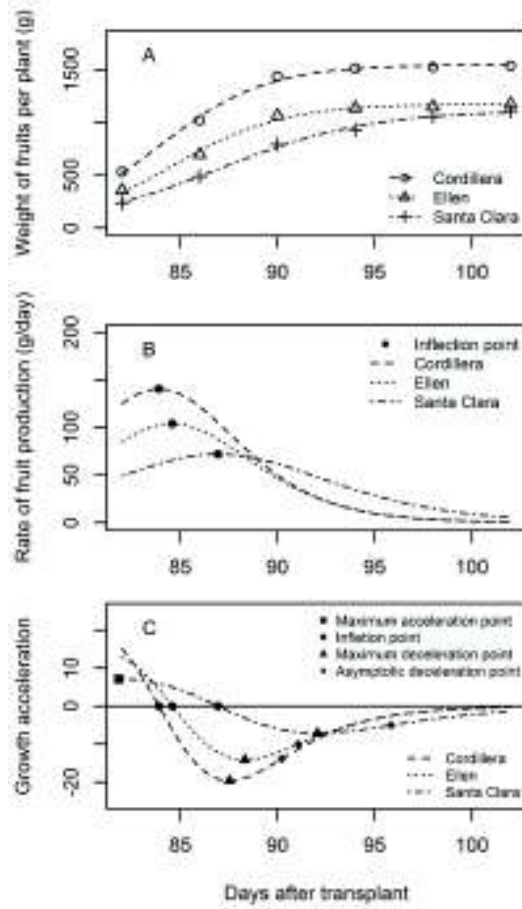
275

276

277

278

279



280

281 **Fig. 4**

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

284

285

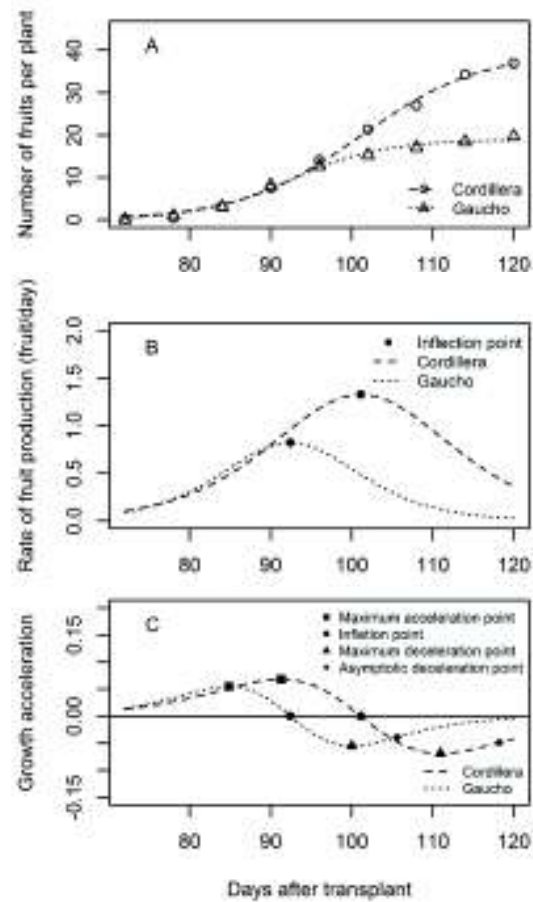
286

287

288

289

290



291

292 **Fig. 5**

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

295

296

297

298

299

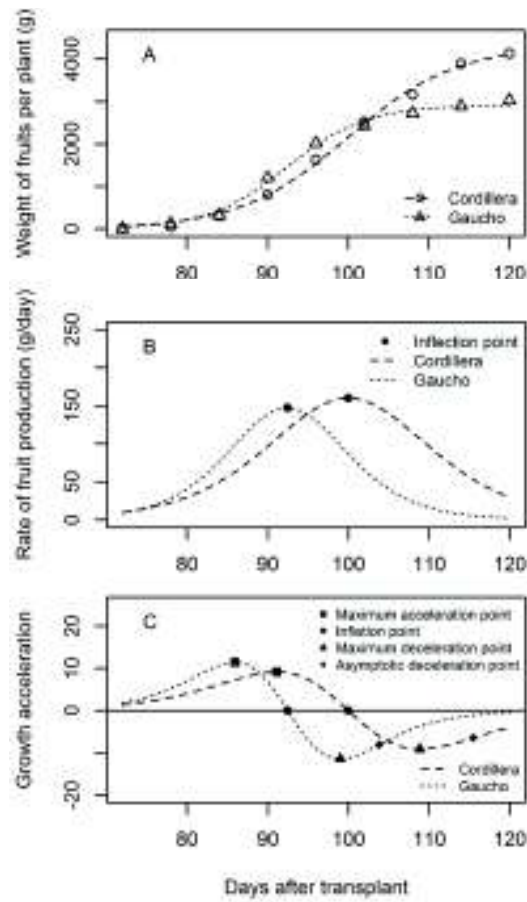
300

301

302

303

304



305

306 **Fig. 6**

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

309

310 The inflection point (IP) occurs when the rate of fruit production is maximal, and from
 311 it the curve changes from an exponential to a logarithmic trend. In the 2015/2016 GS, the IP
 312 was reached approximately 85 DAT in the Cordillera and Ellen genotypes, earlier than in the
 313 Santo Clara genotype, which reached the IP at 87 DAT. In the 2016/2017 GS, the Gaucho
 314 genotype reached this point approximately 92 DAT, while the Cordillera genotype reached
 315 the IP approximately at 100 DAT (Figures 3 to 6).

316 After the IP, the production rate (velocity) tends to decay (decelerate) to a maximum
 317 determined by the maximum deceleration point (MDP). It is observed that, as occurred for the
 318 other critical points, the MDP was reached earlier in the Cordillera and Ellen genotypes in the
 319 2016/2017 GS, and by the Gaucho genotype in the 2016/2017 GS. The asymptotic

320 deceleration point (ADP) indicates the moment of harvest at which increases in production
321 become insignificant. In the 2015/2016 GS, significant increases in yield occurred up to
322 approximately 92 DAT in the Cordillera and Ellen genotypes and up to approximately 97
323 DAT in the Santa Clara genotype. In the 2016/2017 GS the increments were significant up to
324 approximately 106 DAT in the Gaucho genotype, and up to approximately 118 DAT in the
325 Cordillera genotype (Figures 3 to 6).

326

327 **Discussion**

328

329 *Production and climate variables*

330 In the 2015/2016 GS, the Cordillera, Ellen and Santa Clara genotypes produced 51.30,
331 39.17, and 36.95 t ha⁻¹, respectively. The average of tomato production in Rio Grande do Sul
332 state, Brazil, in this GS was 49.08 t ha⁻¹. In the 2016/2017 GS, however, the Cordillera and
333 Gaucho genotypes produced 86.40 and 58.42 t ha⁻¹, respectively, a higher production than the
334 average of production in this GS (51.46 t ha⁻¹) in the State of Rio Grande do Sul, Brazil
335 (IBGE, 2017). The lower production in 2015/2016 GS occurred due to the excess of rainfall
336 during the fruiting, whereas in 2016/2017, the rainfall was centered in the vegetative stage
337 (Figure 7). This situation in addition to the greater number of plants in the 2015/2016 GS than
338 2016/2017 GS, explain this large difference in weight and number of fruits between these two
339 harvests.

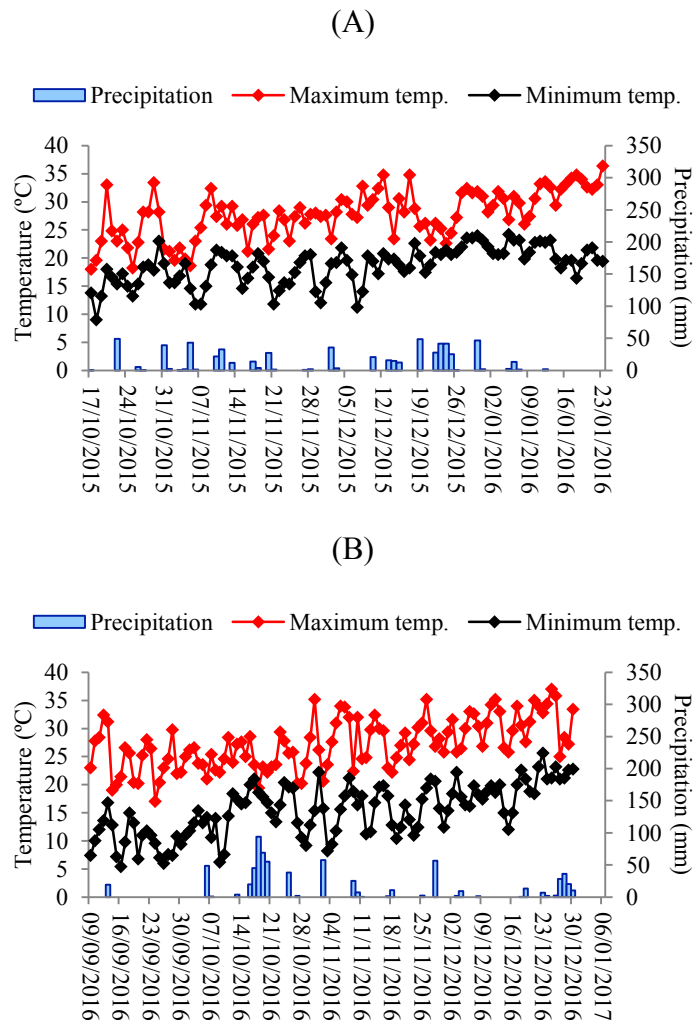
340

341

342

343

344



345 **Fig. 7**
 346 Maximum air temperature, minimum air temperature and daily precipitation in the 2015/16
 347 (A) e 2016/17 (B) GSs.
 348

349 The difference in the productive behavior in the two GS may be related to the
 350 differentiated beginning of the harvests in the two GSs. In the first experiment a later onset of
 351 harvests was observed, with plants producing a high number of fruits (Figures 3 and 4). This
 352 reflects the absence of slower initial growth and the possibility of fitting the Brody model to
 353 the 2015/2016 crop data (Table 2). However, this aspect does not compromise the analysis,
 354 since the lack of adjustment of the Brody model is a result that shows the behavior difference
 355 in the two GSs.

356

357 *Models' fit and assumptions*

358 In relation to the statistical aspects of this study, it is observed that the Bates and Watts
359 nonlinearity has a behavior close to one in logistic model, and that for this reason the
360 parameters of the model are approximately unbiased. In addition, non-violation of the
361 assumptions of normality, homoscedasticity and independence of residuals show that they are
362 estimates of minimum variance, and allow performing hypothesis testing using the F statistic
363 (Seber and Wild, 2003).

364 Even under conditions of heterogeneity and with repeated measures in time, it is
365 verified that the assumptions of homoscedasticity and independence of residuals were not
366 violated (Table 1 and 2). Compliance with homoscedasticity may be related to the use of the
367 grand mean of the replicates in the analysis, but the absence of correlation of the residuals is
368 surprising since the harvests are repeated measures, although this has already been reported in
369 other horticultural crops (Lúcio et al., 2016b). These characteristics, coupled with the high fit
370 of the model and the approximately unbiased parameters estimative, allows for the
371 substitution of the ANOVA with the fitting of growth models to the tomato production data. It
372 also allows for testing of the difference of the parameters of the genotypes through the
373 comparison of nested models using dummy variables (Archontoulis and Miguez, 2015; Ritz
374 and Streibig, 2008).

375 It is emphasized that when the assumption of homoscedasticity of the residuals is
376 violated, the estimation of the parameters must be performed by using the weighted least
377 squares method. When the residuals are heterocedastic and dependent, autoregressive errors
378 should be used, estimating the parameters by the generalized least squares method (Seber and
379 Wild, 2003). Therefore, the nonlinear regression adjustment method itself offers an alternative
380 to situations where there is a violation of the assumptions of the models.

381

382 *Biological interpretation of critical points of the models*

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

390 The scale parameter (β_2) is directly proportional to the difference between the initial
391 production and the asymptote of the curve (in relation of the abscissa) and can therefore be
392 used to determine the degree of maturation (initial production) of the genotypes at the
393 beginning of the harvest. Comparing two genotypes with the same fruit production rate, but
394 different final production, the one with the highest asymptote (final production) will also
395 present a higher estimate of β_2 if the production of the genotypes (same number or mass of
396 fruit) at the beginning of the harvests is the same.

397 It can be seen that even producing a greater mass and number of fruits, the Cordillera
398 genotype presented values of the scale parameter slightly higher than or equal to the other
399 genotypes in the 2015/2016 crop (Figure 1), indicating an early production (higher number or
400 mass of fruit at beginning of the harvest) (Figures 3 and 4). The same occurred in the
401 2016/2017 harvest, when the Cordillera genotype produced a larger number and weight and of
402 fruits, but the value of the scale parameter did not differ statistically from the Gaucho
403 genotype. Although there is no statistically significant difference, the lower value of the
404 estimative of the β_2 indicates the earlier production in the Cordillera genotype. It should be
405 noted that since the harvests were started very early in the second experiment, this difference
406 in initial production was not as high (Figures 5 and 6). Despite the earlier production (higher

407 number or mass of fruit at beginning of the harvest) of the Cordillera genotype in the
408 2015/2016 GS, the high fruit yield (β_3) of the Ellen genotype (statistically equal to the
409 Cordillera genotype) caused both to reach the IP approximately at the same time.

410 The Santa Clara genotype reached the IP later because of the lower fruit production
411 rate, showing itself a late genotype (Figures 3 and 4). In the 2016/2017 GS, the Gaúcho
412 genotype, even showing a late-onset in the production, also reached the IP earlier than the
413 Cordillera genotype due to its higher fruit production rate (Figures 5 and 6). It is observed
414 that, even when it is not statistically different, the magnitude of the fruit production rate
415 interferes significantly in the X_{PI} result, and therefore the isolated analysis of the parameters
416 and critical points is not recommended to determine the precocity of the genotypes.

417 Generally, the absolute values of the number and mass of fruit are related with the β_2
418 value. When the asymptote is higher and the β_2 are relative smaller (in relation to asymptote)
419 the production at beginning of the harvest is greatest. But the moment of occurrence of the PI
420 is more related with higher values of the β_3 .

421 The time in which the other critical points (MAP, MDP and ADP) are reached follow
422 the same behavior of the IP, since they are highly influenced by the fruit production rate
423 (Figures 3 to 6). However, the abscissa values of these points may be important in
424 determining the productive behavior of the genotypes throughout the harvests. For example,
425 one can use the abscissa of these points to delimit the moment in which the production is
426 concentrated. According to Mischan and Pinho (2014), approximately 21.10% of total
427 production (asymptote) occurs when MAP is reached; 50.00% when the IP is reached;
428 78.80% when the MDP is reached; and 90.80% when the ADP is reached.

429 In the 2015/2016 GS, the production of the Cordillera and Ellen genotypes was
430 concentrated in approximately eight days (from 82 to 89 DAT), corresponding to the
431 beginning of the harvest and to the MDP. The remaining 21.20% were produced over 12 days.

432 The production of the Santa Clara genotype included a longer period, concentrating in
433 approximately 10 days, between 83 and 93 DAT, with significant increases in production
434 occurring up to approximately 97 DAT (Figures 3 and 4). Therefore, the Cordillera and Ellen
435 genotypes are characterized by having a more early and concentrated production than the
436 Santa Clara genotype.

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

445 The period between MAP and MDP was short and contained a large part of the
446 production of the studied genotypes. Production before and after these points was small and
447 more dispersed over time. Thus, inferences regarding precocity and crop concentration can
448 also be made using the period between MAP and MDP as a reference. MDP values even
449 indicate until what time production increments are significant. From this moment on, the
450 accumulated production is small (corresponding to less than 10% of the production) and
451 extend for a relatively long period (Figures 3 to 6).

452

453 *Precocity and concentration of production*

454 Keeping constant the other parameters we can determine the precocity by the lower
455 values of the β_2 and higher value the β_3 . But the absolute values of the number or mass of
456 fruit at beginning of the harvest are dependent that the asymptote, as previously reported. It is

457 clear at the Cordillera genotype in the 2015/2016 GS. When the relative value the β_2 is
458 smaller in relation to the asymptote, the initial values of number and mass of fruit are greater.
459 So, we encourage researchers to evaluate the parameters together to determine the precocity
460 of production. In relation at the more concentration of production, it is related with higher β_3 .

461 We proposed determine this characteristic by the moment the occurrence of the PI and
462 the time between MAP and MDP. The time between MAP and MDP also can be used to
463 determine the concentration of production. The earlier occurrence of PI (= more precocity)
464 and the shorter time between MAP and MDP (= more concentration of production) are related
465 with smaller values of the β_2 and higher value of the β_3 . Therefore, we propose (and
466 demonstrate) a clear methodology to determine the precocity and concentration of production,
467 based on the productive behavior of the tomato crop.

468

469 **Conclusions**

470 It has been shown that the use of growth models may be an alternative to ANOVA in
471 statistical analyzes of experiments with multiple harvest horticultural crops. In this study, the
472 problems that could be caused by heteroscedasticity and dependence of residuals were not
473 perceived. However, if they had been observed, this problem could be solved by the
474 application of adequate statistical methods (weighted or generalized least squares). The
475 comparison of the parameters, through the use of nested models, allowed for comparing
476 production and increased the inferential value of the results by determining the initial
477 maturation (through the parameter β_2) and the fruit production rate of the genotypes (through
478 parameter β_3). The critical points allowed for determining the precocity and concentration of
479 the harvests. Using the parameters and critical points of the model, it was possible to
480 determine not only the final total production, but also the maturation of the genotypes at the

481 beginning of the harvest, the productive precocity and the period in which the production is
482 concentrated.

483

484 **Acknowledgement**

485 We thank the Brazilian National Council for Scientific and Technological
486 Development (CNPq) and the Coordination for the Improvement of Higher Level Personnel
487 (CAPES), for granting scholarships to the authors.

488

489 **References**

490 Archontoulis, S. V., Miguez, F.E., 2015. Nonlinear regression models and applications in
491 agricultural research. *Agron. J.* 107, 786–798.

492 Bates, D.M., Watts, D.G., 2007. *Nonlinear Regression Analysis and its Applications*. John
493 Wiley and Sons Inc., New York, pp. 392.

494 Bi, H., Fox, J.C., Li, Y., Lei, Y., Pang, Y., Bi, H., Fox, J., Li, Y., Lei, Y., Pang, Y., 2012.
495 Evaluation of nonlinear equations for predicting diameter from tree height. *Can. J. For.*
496 *Res. J. For. Res* 42, 789–806.

497 Carneiro, A.P.S., Muniz, J.A., Carneiro, P.L.S., Malhado, C.H.M., Filho, R.M., Silva, F.F. e,
498 2014. Pesquisa agropecuária brasileira., *Pesquisa Agropecuária Brasileira*. Embrapa
499 Informação Tecnológica.

500 Goldberg, V., Ravagnolo, O., 2015. Description of the growth curve for Angus pasture-fed
501 cows under extensive systems. *J. Anim. Sci.* 93, 4285. doi:10.2527/jas.2015-9208

502 Helidoniotis, F., Haddon, M., Tuck, G., Tarbath, D., 2011. The relative suitability of the von
503 Bertalanffy, Gompertz and inverse logistic models for describing growth in blacklip
504 abalone populations (*Haliotis rubra*) in Tasmania, Australia. *Fish. Res.* 112, 13–21.
505 doi:10.1016/j.fishres.2011.08.005

- 506 IBGE – Instituto Brasileiro de Geografia e Estatística. Levantamento sistemático da produção
507 agrícola, março de 2017, 2017. Instituto Brasileiro de Geografia e Estatística, Rio de
508 Janeiro, pp. 83.
- 509 Katsanevakis, S., Maravelias, C.D., 2008. Modelling fish growth: Multi-model inference as a
510 better alternative to a priori using von Bertalanffy equation. *Fish Fish.* 9, 178–187.
511 doi:10.1111/j.1467-2979.2008.00279.x
- 512 Leduc, D., Goelz, J., 2009. A Height-Diameter Curve for Longleaf Pine Plantations in the
513 Gulf Coastal Plain. *South. J. Appl. For.* 33, 164–170.
- 514 Lúcio, A.D., Sari, B.G., Pezzini, R. V, Liberalesso, V., Delatorre, F., Faé, M. 2016.
515 Heterocedasticidade entre fileiras e colheitas de caracteres produtivos de tomate cereja e
516 estimativa do tamanho de parcela. *Hortic. Bras.* 34, 223–230. doi:10.1590/S0102-
517 053620160000200012
- 518 Lúcio, A.D., Sari, B.G., Rodrigues, M., Bevilaqua, L.M., Voss, H.M.G., Copetti, D., Faé, M.,
519 2016. Modelos não-lineares para a estimativa da produção de tomate do tipo cereja.
520 *Ciência Rural* 46, 233–241. doi:10.1590/0103-8478cr20150067
- 521 Lúcio, A.D.C., Couto, M.R.M., Lopes, S.J., Storck, L., 2011. Transformação box-cox em
522 experimentos com pimentão em ambiente protegido. *Hortic. Bras.* 29, 38–42.
523 doi:10.1590/S0102-05362011000100007
- 524 Lúcio, A.D.C., Nunes, L.F., Rego, F., 2015. Nonlinear models to describe production of fruit
525 in Cucurbita pepo and Capsicum annum. *Sci. Hortic. (Amsterdam)*. 193, 286–293.
526 doi:10.1016/j.scienta.2015.07.021
- 527 Mehtätalo, L., De-Miguel, S., Gregoire, T.G., 2015. Modeling height-diameter curves for
528 prediction. *Can. J. For. Res.* 45, 826–837. doi:10.1139/cjfr-2015-0054
- 529 Mendonça, A.R. de, Calegario, N., Silva, G.F. da, Souza, A.L. de, Trugilho, P.F., Carvalho,
530 S.P.C. e, Possato, E.L., 2014. Modelagem da produção de sortimentos em povoamentos

- 531 de eucalipto. CERNE 20, 587–594. doi:10.1590/01047760201420041366
- 532 Mischan, M.M., de Pinho, S.Z., de Carvalho, L.R., 2011. Determination of a point sufficiently
533 close to the asymptote in nonlinear growth functions. *Sci. Agric.* 68, 109–114.
534 doi:10.1590/S0103-90162011000100016
- 535 Mischan, M.M., Pinho, S.Z., 2014. Modelos não lineares: funções assintóticas de
536 crescimento. *Cultura Acadêmica*, São Paulo, pp. 184.
- 537 Pinho, S.Z. de, Carvalho, L.R. de, Mischan, M.M., Passos, J.R. de S., 2014. Critical points on
538 growth curves in autoregressive and mixed models. *Sci. Agric.* 71, 30–37.
539 doi:10.1590/S0103-90162014000100004
- 540 Ritz, C., Streibig, J.C., 2008. *Nonlinear Regression with R*. Springer, New York, pp. 142.
- 541 Regazzi, A.J., 2003. Teste para verificar a igualdade de parâmetros e a identidade de modelos
542 de regressão não linear. *Ceres*, 50, 9-26.
- 543 Seber, G.A.F., Wild, C.J., 2003. *Nonlinear regression*. John Wiley and Sons Inc., New Jersey,
544 pp. 768.
- 545 Wubs, A.M., Ma, Y.T., Heuvelink, E., Hemerik, L., Marcelis, L.F.M., 2012. Model Selection
546 for Nondestructive Quantification of Fruit Growth in Pepper. *J. Amer. Soc. Hort. Sci.*
547 137, 71–79.
- 548

5 DISCUSSÃO GERAL

Todos os genótipos de tomate testados apresentaram variáveis produtivas (massa e número de frutos) acumulada com comportamento sigmoide, o que permitiu a utilização dos modelos de crescimento (SEBER; WILD, 2003). Entre os modelos testados, o logístico foi utilizado, pois além de apresentar boa qualidade de ajuste também foi o que apresentou menor não linearidade paramétrica. A não linearidade intrínseca e paramétrica baixa (valores menores que 0,3 e 1, respectivamente) é extremamente importante, pois elas estão relacionadas com estimativas aproximadamente não viesadas dos parâmetros (BATES; WATTS, 1988; SEBER; WILD, 2003). No caso dos genótipos testados, todos os pressupostos foram cumpridos, e testes de hipóteses foram realizados para verificar a igualdade dos parâmetros entre os genótipos (utilizando o teste F).

A seleção de modelos é uma etapa crucial com quem utiliza os modelos não lineares. A estimação parâmetros aproximadamente não viesados deve ser um critério de seleção de modelos (FERNANDES et al., 2015), porém a seleção é realizada (na maioria das vezes) apenas com base apenas na qualidade de ajuste (maior R^2 , menor AIC, etc), o que pode levar a resultados equivocados. Como é a interpretação dos parâmetros que possibilitam o aumento das inferências sobre o comportamento produtivo das culturas olerícolas ao longo do tempo, a estimação de parâmetros aproximadamente não viesados é fundamental para que eles descrevam a realidade observada. Por isso, a aproximação linear dos modelos deve ser avaliada utilizando a) as curvaturas de Bates e Watts, b) os gráficos de contornos (BATES; WATTS, 1988; SEBER; WILD, 2003), c) medidas de Hougaard (CARVALHO et al., 2010), d) bias das estimativas (utilizando a metodologia de Box (1971) ou reamosntragens “*bootstrap*”) ou e) normalidade da distribuição empírica das estimativas dos parâmetros obtidos por “*bootstrap*” (Seber and Wild, 2003; Ritz and Streibig, 2008).

As pressuposições de normalidade homocedasticidade e independência dos resíduos são essenciais para que o processo inferencial seja realizado. Esses pressupostos estão relacionados com a precisão das estimativas, e a sua violação levam a estimativas do intervalo de confiança dos parâmetros pouco precisa ou a inviabilidade do teste de hipóteses para parâmetros (MUNIZ et al., 2017). Porém, o descumprimento dos pressupostos não impede o uso dos modelos não lineares, pois esse problema pode ser contornado de duas formas:

a) estimação dos parâmetros utilizando o método dos mínimos quadrados ponderados e generalizados, quando os erros são heterocedásticos ou dependentes, respectivamente;

b) estimação dos intervalos por reamostragens “*bootstrap*”, que permite construir intervalos de confiança.

Além dos parâmetros, os pontos críticos do modelo logístico (que foi o selecionado) também foram calculados, e os seus valores foram associados com as seguintes características:

a) Produção total: corresponde ao valor de β_1 ;

b) Precocidade produtiva: está associada a valores baixos de β_2 e elevados de β_3 (pontos críticos ocorrem mais cedo).

c) Concentração da produção: está associado a valores elevados de β_3 (diminui o tempo entre o PDA e o PDM, associado ao crescimento exponencial da produção).

Essas características foram abordadas de forma distinta nos dois artigos da tese. Outra forma de abordagem poderia ser a utilização dos parâmetros e dos pontos críticos do modelo como variáveis explicativas em um modelo da ANOVA, com testes complementares (de médias) realizados posteriormente. De qualquer maneira, os cuidados com a estimação dos parâmetros (principalmente em relação a não linearidade) continuariam tendo importância fundamental.

Quanto a produção, precocidade produtiva e concentração de produção dos genótipos avaliados nos dois anos de cultivo, foi observado o seguinte cenário (Tabelas 1 e 2):

Tabela 10- Número médio de frutos por planta e concentração de produção de frutos estimados pelo modelo logístico em cada genótipo de tomate nos dois anos de condução dos experimentos.

Genótipo	Ano	Produção ¹	Pico de produção (DAT) ²	Concentração da produção (DAT) ³
Cordillera		17,58	85	82 a 89
Ellen	2015/2016	8,84	86	82 a 90
Santa Clara		13,35	88	83 a 93
Cordillera	2016/2017	39,68	101	92 a 111
Gaucho		18,94	93	85 a 101

¹ Estimado pelo valor de β_1 ; ² Estimado pela abcissa do ponto de inflexão; ³ Estimado pelas abcissas dos pontos de aceleração máxima e desaceleração máxima.

Tabela 2 - Massa média de frutos por planta (g) e concentração de produção de frutos estimados pelo modelo logístico em cada genótipo de tomate nos dois anos de condução dos experimentos.

Genótipo	Ano	Produção ¹	Pico de produção (DAT) ²	Concentração da produção (DAT) ³
Cordillera		1549,72	84	81 a 88
Ellen	2015/2016	1178,20	85	81 a 88
Santa Clara		1118,52	87	82 a 93
Cordillera	2016/2017	4320,11	100	92 a 109
Gaucho		2921,10	93	86 a 99

¹ Estimado pelo valor de β_I ; ² Estimado pela abcissa do ponto de inflexão; ³ Estimado pelas abcissas dos pontos de aceleração máxima e desaceleração máxima.

No primeiro ano de cultivo, o genótipo Cordillera foi o mais produtivo e precoce. A produção do genótipo Ellen foi menor, porém se concentrou no mesmo período que no genótipo Cordillera. Por fim, o genótipo Santa Clara caracterizou-se por ser o menos precoce e com produção menos concentrada. No segundo ano o genótipo Gaúcho concentrou sua produção em um menor período de tempo em relação ao genótipo Cordillera, podendo isso estar relacionado com a baixa produção de frutos.

Ressalta-se que não é papel deste trabalho encontrar explicação fisiológica para os fatos relatados, mas sim demonstrar o potencial da utilização dos modelos de crescimento como ferramenta de análise estatística. Executando de forma correta os procedimentos de ajuste e seleção de modelos, os pesquisadores poderão realizar inferências pormenorizadas do comportamento produtivo de genótipos.

Apesar do trabalho ter sido desenvolvido com a cultura do tomate, ele também pode ser utilizado com qualquer cultura olerícola de múltiplas colheitas. Portanto, a utilização dos modelos de crescimento é uma alternativa a ANOVA, e tem como vantagem contornar os problemas de violação dos pressupostos (heterocedasticidade e a dependência dos resíduos) e aumentar as inferências que podem ser realizadas quanto ao comportamento produtivo das culturas olerícolas de múltiplas colheitas.

6 CONCLUSÕES

- As variáveis massa e número de frutos de tomate (acumuladas ao longo das colheitas) tem um comportamento sigmoide;
- Entre os modelos testados, o logístico é o que melhor descreve o comportamento destas variáveis;
- Pelas estimativas dos parâmetros e dos pontos críticos do modelo logístico é possível fazer inferências quanto a produção total (de massa e número de frutos), precocidade produtiva e concentração de produção do tomate;
- Os modelos de crescimento podem ser utilizados como alternativa a ANOVA na análise de experimentos com tomate.
- O genótipo Cordillera foi o mais precoce e produtivo no primeiro ano de cultivo. No segundo ano, o genótipo Gaucho foi o menos produtivo, porém o mais precoce.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v.19, n.6, p. 717–723, 1974. Disponível em: <https://link.springer.com/chapter/10.1007/978-1-4612-1694-0_16>. Acesso em: 22 jan. 2018. DOI: 10.1007/978-1-4612-1694-0_16.

ARCHONTOULIS, S.V.; MIGUEZ, F.E. Nonlinear Regression Models and Applications in Agricultural Research. **Agonomy Journal**, v. 107, n. 2, p. 786-798, 2015. Disponível em: <<https://dl.sciencesocieties.org/publications/aj/articles/107/2/786>>. Acesso em: 22 jan. 2018. DOI: 10.2134/agronj2012.0506.

BATES, D.M.; WATTS, D.G. **Nonlinear regression and its applications**. New York: John Wiley & Sons, 1988. 365 p.

BATY, F et al. A toolbox for nonlinear regression in R: the package nlstools. **Journal of Statistical Software**, v.66, n.5, p.1-21, 2015. Disponível em: <<https://www.jstatsoft.org/article/view/v066i05>>. Acesso em: 22 jan. 2018. DOI: 10.18637/jss.v066.i05.

BEALE, E.M.L. Confidence regions in non-linear estimation. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 22, n.1, p.41–88, 1960. Disponível em: <https://www.jstor.org/stable/2983877?seq=1#page_scan_tab_contents>. Acesso em: 22 jan. 2018.

BOX, M.J. Bias in Nonlinear Estimation. **Journal of the Royal Statistical Society. Series B (Methodological)**, v33, p.171-201, 1971. Disponível em: <<http://www.jstor.org/stable/2985002>>. Acesso em: 09 abr. 2018.

CARGNELUTTI FILHO, A. et al. Número de colheitas e comparação de genótipos de tomateiro cultivados em estufa de plástico. **Pesquisa Agropecuária Brasileira**, v.39, n.10, p.953-959, 2004. Disponível em: <<http://www.scielo.br/pdf/pab/v39n10/22316.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0100-204X2004001000002.

CARVALHO, L.R. et al. Methods to verify parameter equality in nonlinear regression models. **Scientia Agricola**, v.67, n.2, p.218-222, 2010. Disponível em: <<http://www.scielo.br/pdf/sa/v67n2/a14v67n2.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0103-90162010000200014.

CASELLA, G. **Statistical design**. New York: Springer, 2008. 307 p.

COUTO, M.R.M. et al. Transformações de dados em experimentos com abobrinha italiana em ambiente protegido. **Ciência Rural**, v.39, n.6, p.1701-1707, 2009. Disponível em: <<http://www.scielo.br/pdf/cr/v39n6/a233cr1169.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0103-84782009005000110.

DRAPER, N.R.; SMITH, H. **Applied regression analysis**. 3 ed. New York: John Wiley & Sons, 1998. 706 p.

ELLI, E.P. et al. Precision of growth estimates and sufficient sample size: can solar radiation level change these factors? **Agronomy Journal**, v.110, n. 1, p.155–163, 2018. Disponível em: <<https://dl.sciencesocieties.org/publications/aj/abstracts/110/1/155?access=0&view=pdf>> Acesso em: 22 jan. 2018. DOI: 10.2134/agronj2017.05.0297.

EMILIANO, P.C. et al. Information criteria: How do they behave in different models? **Computational Statistics and Data Analysis**, v.69, p.141–153, 2014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167947313002776>>. Acesso em: 22 jan. 2018. DOI: 10.1016/j.csda.2013.07.032.

FERNANDES, T.J. et al. Parameterization effects in nonlinear models to describe growth curves. **Acta Scientiarum. Technology**, v. 37, n. 4, p. 397-402, 2015. Disponível em: <<http://periodicos.uem.br/ojs/index.php/ActaSciTechnol/article/view/27855>>. Acesso em: 22 jan. 2018. DOI: 10.4025/actascitechnol.v37i4.27855.

FERNANDEZ, G.C.J. Design and analysis of commonly used comparative horticultural experiments. **Hortscience**, v. 42, n.5, p.1052-1069, 2007. Disponível em: <<http://hortsci.ashspublications.org/content/42/5/1052.full>>. Acesso em: 22 jan. 2018.

FOX, J. **An R and S-Plus companion to applied regression**. London: Sage, 2002. 312 p.

FOX, J.; WEISBERG, S. **An R and S-Plus companion to applied regression**. 2 ed. London: Sage, 2011. 449 p.

GROSS, J.; LIGGES, U. **Package “nortest”**. 2015. Disponível em: <<https://cran.r-project.org/web/packages/nortest/nortest.pdf>>. Acesso em: 22 jan. 2018.

HOTHORN, T. et al. **Package “lmtest”**. 2017. Disponível em: <<https://cran.r-project.org/web/packages/lmtest/lmtest.pdf>>. Acesso em: 22 jan. 2018.

KUTNER, M.H. et al. **Applied linear statistical models**. 5 ed. New York: McGraw-Hill/Irwin, 2004. 1396 p.

LUCIO, A.D.; SARI, B.G. Planning and implementing experiments and analyzing experimental data in vegetable crops: problems and solutions. **Horticultura Brasileira**, v.35, n.3., p. 316-327, 2017. Disponível em: <<http://www.scielo.br/pdf/hb/v35n3/1806-9991-hb-35-03-00316.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/s0102-053620170302.

LÚCIO, A.D.C. et al. Nonlinear models to describe production of fruit in *Cucurbita pepo* and *Capiscum annuum*. **Scientia Horticulturae**, v.193, p.286-293, 2015. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304423815300960>>. Acesso em: 22 jan. 2018. DOI: 10.1016/j.scienta.2015.07.021.

LÚCIO, A.D.C. et al. Nonlinear regression and plot size to estimate green beans production. **Horticultura Brasileira**, v34, n. 4, p.507-513, 2016a. Disponível em: <<http://www.scielo.br/pdf/hb/v34n4/1806-9991-hb-34-04-00507.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/s0102-053620160409.

LÚCIO, A.D.C. et al. Heterocedasticidade entre fileiras e colheitas de caracteres produtivos de tomate cereja e estimativa do tamanho de parcela. **Horticultura Brasileira**, v. 34, n. 2, p. 223-230, 2016. Disponível em: <<http://www.scielo.br/pdf/hb/v34n2/1806-9991-hb-34-02-00223.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0102-053620160000200012.

LÚCIO, A.D.C. et al. Modelos não-lineares para a estimativa da produção de tomate do tipo cereja. **Ciência Rural**, v.46, n.2, p.233-241, 2016b. Disponível em: <http://www.scielo.br/pdf/cr/v46n2/1678-4596-cr-0103_8478cr20150067.pdf>. Acesso em: 22 jan. 2018. DOI: 10.1590/0103-8478cr20150067.

LÚCIO, A.D.C. et al. Violação dos pressupostos do modelo matemático e transformação de dados. **Horticultura Brasileira**, v. 30, n. 3, p. 415-423, 2012. Disponível em: <<http://www.scielo.br/pdf/hb/v30n3/10.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0102-05362012000300010.

LÚCIO, A.D.C. et al. Excesso de zeros nas variáveis observadas: estudo de caso em experimento com brócolis. **Bragantia**, v. 69, n. 4, p.1035-1046, 2010. Disponível em: <<http://www.scielo.br/pdf/brag/v69n4/v69n4a32.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0006-87052010000400032.

MISCHAN, M.M. et al. Inflection and stability points of diphasic logistic analysis of growth. **Scientia Agricola**, v.72, n.3, p.215-220, 2015. Disponível em: <<http://www.scielo.br/pdf/sa/v72n3/0103-9016-sa-72-3-0215.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/0103-9016-2014-0212.

MISCHAN, M.M.; PINHO, S.Z. **Modelos não lineares: funções assintóticas de crescimento**. São Paulo: Cultura Acadêmica, 2014. 184 p.

MISCHAN, M.M. et al. Determination of a point sufficiently close to the asymptote in nonlinear growth functions. **Scientia Agricola**, v.68, n.1, p.109-114, 2011. Disponível em: <<http://www.scielo.br/pdf/sa/v68n1/v68n1a16.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0103-90162011000100016.

MUNIZ, J.A. et al. Nonlinear models for description of cacao fruit growth with assumption violations. **Revista Caatinga**, v. 30, n. 1, p. 250 – 257, 2017. Disponível em: <<http://www.scielo.br/pdf/rcaat/v30n1/1983-2125-rcaat-30-01-00250.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/1983-21252017v30n128rc.

OLIVOTO, T. et al. Photosynthetic photon flux density levels affect morphology and bromatology in *Cichorium endivia* L. var. *latifolia* grown in a hydroponic system. **Scientia Horticulturae**, v.230, p.178–185. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304423817307124>>. Acesso em: 22 jan. 2018. DOI:10.1016/j.scienta.2017.11.031.

PAINE, C.E.T. et al. How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. **Methods in Ecology and Evolution**, v.3, n.2, p.245–256, 2012. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2011.00155.x/abstract>>. Acesso em: 22 jan. 2018. DOI: 10.1111/j.2041-210X.2011.00155.x.

PASSOS, J.R.S. et al. Critical points in logistic growth curves and treatment comparisons. **Scientia Agricola**, v.69, n.5, p.308-312, 2012. Disponível em: <<http://www.scielo.br/pdf/sa/v69n5/a04v69n5.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0103-90162012000500004.

PINHEIRO J. et al. **Package “nlme”**. 2017. Disponível em: <<https://cran.r-project.org/web/packages/nlme/nlme.pdf>>. Acesso em: 22 jan. 2018.

PINHO, S.Z. et al. Critical points on growth curves in autoregressive and mixed models. **Scientia Agricola**, v.71, n.1, p.30-37, 2014. Disponível em: <<http://www.scielo.br/pdf/sa/v71n1/v71n1a04.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0103-90162014000100004.

R Core Team (2017). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria.

RATKOWSKY, D.A. Principles of nonlinear regression modeling. **Journal of Industrial Microbiology**, v.12, n.3-5, p. 195-199, 1993. Disponível em: <<https://link.springer.com/article/10.1007/BF01584190>>. Acesso em: 22 jan. 2018. DOI: 10.1007/BF01584190.

REGAZZI, A.J.; SILVA, C.H.O. Testes para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear em dados de experimento com delineamento em blocos casualizados. **Revista Ceres**, v. 57, n.3, p. 315-320, 2010. Disponível em: <<http://www.scielo.br/pdf/rceres/v57n3/v57n3a05.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0034-737X2010000300005.

RENCHER, A.C.; SCHAALJE, G.B. **Linear models in statistics**. 2 ed. New Jersey: John Wiley & Sons, 2008. 672 p.

RIPLEY, B.; VENABLES, B. **Modern applied statistics with S**. 4 ed. New York: Springer, 2002. 495 p.

RIPLEY, B. et al. **Package ‘MASS’**. 2017. Disponível em: <<https://cran.r-project.org/web/packages/MASS/MASS.pdf>>. Acesso em: 22 jan. 2018.

RITZ, C. **Package “nlrwr”**. 2013. Disponível em: <<https://cran.r-project.org/src/contrib/Archive/nlrwr>>. Acesso em: 22 jan. 2018.

RITZ, C.; STREIBIG, J.C. **Nonlinear Regression with R**. New York: Springer, 2008. 114 p.

SAVIAN, T.V.; MUNIZ, J.A. Study of *in situ* degradability: heterogeneity of variances and correlated errors. **Scientia Agricola**, v.64, n.5, p.548-554, 2007. Disponível em: <<http://www.scielo.br/pdf/sa/v64n5/v64n5a13.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/S0103-90162007000500013.

SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**, v.6, n.2, p.461–464, 1978. Disponível em: <<https://projecteuclid.org/euclid.aos/1176344136>>. Acesso em: 22 jan. 2018.

SEBER, G.A.F.; WILD, C.J. **Nonlinear regression**. New York: John Wiley & Sons, 2003. 706 p. 768 p.

SOUSA, I.F.; et al. Fitting nonlinear autoregressive models to describe coffee seed germination. **Ciência Rural**, v.44, n.11, p.2016-2021, 2014. Disponível em:

<<http://www.scielo.br/pdf/cr/v44n11/0103-8478-cr-44-11-02016.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.1590/0103-8478cr20131341.

SUGIURA, N. Further analysts of the data by Akaike's information criterion and the finite corrections. **Communications in Statistics - Theory and Methods**, v.7, n.1, p. 13–26, 1978. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/03610927808827599>>. Acesso em: 22 jan. 2018. DOI: 10.1080/03610927808827599.

UEDA, C.M. et al. Nonlinear models for describing the citrus variegated chlorosis in groves of two counties at northwestern Paraná state, Brazil. **Acta Scientiarum. Agronomy**, v. 32, n. 4, p. 603-611, 2010. Disponível em: <<http://www.scielo.br/pdf/asagr/v32n4/a06v32n4.pdf>>. Acesso em: 22 jan. 2018. DOI: 10.4025/actasciagron.v32i4.11625.

ZEVIANI, et al. Modelos não lineares para a liberação de potássio de esterco animal em latossolos. **Ciência Rural**, v.42, n.10, p.1789-1796, 2012. Disponível em: <<http://www.scielo.br/pdf/cr/v42n10/a28712cr3006.pdf>>. Acesso em: 09 abr. 2018. DOI: 10.1590/S0103-84782012001000012.

ANEXO A – CARACTERÍSTICAS DOS GENÓTIPOS DE TOMATE

Genótipo	Tipo	Hábito de crescimento	Peso dos frutos	Características gerais
Cordillera	Saladete	Indeterminado	140 – 160 g	- Consumo in natura - Tolerante à doenças - Alta produção
Ellen	Salada	Indeterminado	200 – 260 g	- Consumo in natura - Tolerante à doenças - Alta produção
Santa Clara	Santa Cruz	Indeterminado	130 – 150 g	- Consumo in natura - Uniformidade dos frutos
Gaúcho	Salada	Indeterminado	220 – 250 g	- Consumo in natura - Uniformidade dos frutos - Excelente sabor - Suscetibilidade a doenças