Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Modelo não linear Chanter: uma aplicação aos dados de
crescimento de frutos do cacaueiro

Pollyane Vieira da Silva

Dissertação apresentada para obtenção do título
de Mestra em Ciências. Área de concentração:
Estatística e Experimentação Agronômica

Piracicaba
2018

Pollyane Vieira da Silva
Licenciada em Matemática

**Modelo não linear Chanter: uma aplicação aos dados de crescimento de frutos
do cacaueiro**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:
Prof ^a Dr^a **TACIANA VILLELA SAVIAN**

Dissertação apresentada para obtenção do título de Mestra
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

**Piracicaba
2018**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Silva, Pollyane Vieira da
Modelo não linear Chanter: uma aplicação aos dados de crescimento de
frutos do cacauzeiro / Pollyane Vieira da Silva. – versão revisada de acordo com
a resolução CoPGr 6018 de 2011. – Piracicaba, 2018.
81 p.

Dissertação (Mestrado) – USP / Escola Superior de Agricultura “Luiz
de Queiroz”.

. 1. Medidas de não linearidade 2. Modelos de crescimento 3. Regressão
não linear I. Título.

AGRADECIMENTOS

Primeiramente agradeço minha família, meu pai Paulo e minha mãe Eliane que sempre estiveram presentes me apoiando, meus irmãos Paulinho e Gabriela, por todas as críticas construtivas. Sempre me senti segura para continuar com vocês ao meu lado.

Um agradecimento especial ao meu namorado Givanildo, por toda ajuda, companheirismo e parceria de sempre. Sou uma pessoa melhor depois que te conheci.

Agradeço a minha orientadora Taciana, por gentilmente ter me ajudado e me guiado no decorrer deste trabalho, me dando todo o suporte necessário. Tenho muita sorte de encontrar pessoas boas no meu caminho, e você é uma delas.

Agradeço também aos meus amigos e colegas da ESALQ por toda a ajuda, aos meus queridos alunos, aos amigos da escola Pedro Raphael, aos amigos do CEPRA e amigos do Bayeux por toda torcida e apoio.

Ao Departamento de Ciências Exatas - LCE, em especial aos professores com quem tive a oportunidade de ter aulas, obrigada por todo conhecimento que adquiri com vocês, agradeço também aos funcionários do departamento, em especial a Solange, por toda ajuda e informação. Ao apoio financeiro da CAPES para o desenvolvimento e realização desse trabalho.

Agradeço a Deus por mais essa vitória!

SUMÁRIO

RESUMO	5
ABSTRACT	6
LISTA DE FIGURAS	7
LISTA DE TABELAS	10
1 INTRODUÇÃO	11
2 REVISÃO BIBLIOGRÁFICA	12
2.1 Regressão	12
2.2 Regressão não linear	14
2.3 Método iterativo de Gauss-Newton	16
2.4 Modelos de crescimento	17
2.4.1 Modelo Logístico	18
2.4.2 Modelo Gompertz	21
2.4.3 Modelo Chanter	24
2.5 Critérios para qualidade do ajuste	25
3 MATERIAL E MÉTODOS	30
3.1 Material	30
3.1.1 Conjunto de dados	30
3.1.2 Modelos	30
3.2 Métodos	31
3.2.1 Simulação	31
3.2.2 Ajuste empírico dos modelos	32
4 RESULTADOS E DISCUSSÃO	34
4.1 Estudo da função Chanter	34
4.2 Estudo de simulação	44
4.3 Estudo dos dados reais	56
5 CONCLUSÃO	65
REFERÊNCIAS	66
APÊNDICES	68
ANEXOS	72

RESUMO

Modelo não linear Chanter: uma aplicação aos dados de crescimento de frutos do cacauzeiro

Modelos não lineares como o Logístico e o Gompertz são amplamente usados para descrever vários processos biológicos por meio da curva de crescimento dada pela equação do modelo. O objetivo deste trabalho foi ajustar o modelo Chanter, assim como o Logístico e o Gompertz, utilizando um conjunto de dados do fruto do cacauzeiro. O modelo Chanter é um híbrido entre o modelo Logístico e o modelo Gompertz cujos parâmetros podem ser interpretados similarmente. A comparação sobre a qualidade do ajuste entre os modelos foi feita utilizando as seguintes medidas estatísticas: o critério de informação de Akaike (AIC), o critério Peso de Akaike, o critério de informação de Bayes (BIC), o desvio padrão residual (DPR) e as medidas de não linearidade vício de Box e curvatura de Bates e Watts além de um estudo de simulação. Verificou-se que o modelo Chanter dentre os modelos estudados neste trabalho é o mais adequado para o ajuste dos dados do fruto do cacauzeiro.

Palavras-chave: Medidas de não linearidade, Modelos de crescimento e Regressão não linear.

ABSTRACT

Chanter Nonlinear Model: an application to cocoa fruits growth data

Nonlinear models such as Logistic and Gompertz are widely used to describe several biological processes using a growth curve given by the equation of the model. The objective of this work was to adjust the Chanter model, as well as the Logistic and the Gompertz, using a data set of cocoa fruit. The Chanter model is a hybrid between the Logistic model and the Gompertz model whose parameters can be interpreted similarly. A comparison of the quality of fit between the models was made using the following statistical measures: the Akaike information criterion (AIC), the Akaike weight criterion, Bayes information criterion (BIC), residual standard deviation (RSD), and measures of non-linearity Box addition and Bates and Watts curvature as well as a simulation study. It was verified that the Chanter model is the most suitable one among the studied models for modeling the cocoa data.

Keywords: Growth models, Measures of non-linearity e Nonlinear regression.

LISTA DE FIGURAS

Figura 1 - Gráficos da função Logística para diferentes valores dos parâmetros β_1 sendo (a) $\beta_2 = -1$ e $\beta_3 = 2$ e (b) $\beta_2 = 1$ e $\beta_3 = 2$	19
Figura 2 - Gráficos da função Logística para diferentes valores dos parâmetros β_2 sendo $\beta_1 = 2$ e $\beta_3 = 2$	20
Figura 3 - Gráficos da função Logística para diferentes valores dos parâmetros β_3 sendo (a) $\beta_1 = 3$ e $\beta_2 = -1$ e (b) $\beta_1 = 3$ e $\beta_2 = 1$	20
Figura 4 - Representações gráficas da função Gompertz sendo (a) $\beta_2 = 1$ e $\beta_3 = 2$ fixos e β_1 variando, (b) $\beta_1 = 4$ e $\beta_3 = 2$ fixos e β_2 variando e (c) $\beta_1 = 3$ e $\beta_2 = 1$ fixos e β_3 variando	23
Figura 5 - Representação gráfica da função Chanter para $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 0,5$ e $\beta_4 = -1$	24
Figura 6 - Representação gráfica da relação entre a diferença nos valores de AICc e a probabilidade de que cada modelo seja verdadeiro (Fonte: MOTULSKY; CHRISTOPOULOS, 2003, p.145)	26
Figura 7 - Representações gráficas da função Chanter para $\beta_1 = 2$, $\beta_2 = 3$ e $\beta_3 = 1$, (a) $\beta_4 = 0,5$ e (b) $\beta_4 = -0,5$	35
Figura 8 - Gráficos da função Chanter para diferentes valores do parâmetro β_1 sendo $\beta_2 = 3$ e $\beta_3 = 1$ (a) $\beta_4 = 0,5$ e (b) $\beta_4 = -0,5$	37
Figura 9 - Gráficos da função Chanter para diferentes valores do parâmetro β_3 sendo $\beta_1 = 2$, $\beta_2 = 3$ (a) $\beta_4 = 0,5$ e (b) $\beta_4 = -0,5$	38
Figura 10 - Gráficos da função Chanter para $\beta_1 = 3$, $\beta_2 = 4$ e $\beta_3 = 2$ e diferentes valores de β_4	39
Figura 11 - Cenário 1: Histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de comprimento do fruto em que linha em vermelho indica o valor paramétrico e a linha em vermelho tracejada indica a média das estimativas	44
Figura 12 - Cenário 1: Box-plot dos vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de comprimento do fruto	45
Figura 13 - Cenário 1: Gráficos de dispersão dos AIC da simulação com os vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidos a partir dos dados simulados de comprimento do fruto	46

Figura 14 - Cenário 1: Histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de diâmetro do fruto em que linha em vermelho indica o valor paramétrico e a linha em vermelho tracejada indica a média das estimativas	47
Figura 15 - Cenário 1: Box-plot dos vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de diâmetro do fruto	48
Figura 16 - Cenário 1: Gráficos de dispersão dos AIC da simulação com os vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidos a partir dos dados simulados de diâmetro do fruto	49
Figura 17 - Cenário 2: Histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de comprimento do fruto em que linha em vermelho indica o valor paramétrico e a linha em vermelho tracejada indica a média das estimativas	50
Figura 18 - Cenário 2: Box-plot dos vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de comprimento do fruto	51
Figura 19 - Cenário 2: Gráficos de dispersão dos AIC da simulação com os vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidos a partir dos dados simulados de comprimento do fruto	52
Figura 20 - Cenário 2: Histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de diâmetro do fruto em que linha em vermelho indica o valor paramétrico e a linha em vermelho tracejada indica a média das estimativas	53
Figura 21 - Cenário 2: Box-plot dos vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de diâmetro do fruto	54
Figura 22 - Cenário 2: Gráficos de dispersão dos AIC da simulação com os vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidos a partir dos dados simulados de diâmetro do fruto	55
Figura 23 - Gráficos de dispersão dos dados de comprimento e diâmetro de cacau ao longo do tempo	56
Figura 24 - Médias e ajuste dos modelos Logístico, Gompertz e Chanter aos dados de: (a) comprimento e (b) diâmetro de cacau ao longo do tempo	57

Figura 25 - Gráficos dos resíduos padronizados para a medida de comprimento do cacau para os modelos Logístico, Gompertz e Chanter respectivamente . .	59
Figura 26 - Gráficos dos resíduos padronizados para a medida de diâmetro do cacau para os modelos Logístico, Gompertz e Chanter respectivamente	59
Figura 27 - Gráficos Normal de probabilidade para a medida de comprimento do cacau para os modelos Logístico, Gompertz e Chanter respectivamente	59
Figura 28 - Gráficos Normal de probabilidade para a medida de diâmetro do cacau para os modelos Logístico, Gompertz e Chanter respectivamente	60
Figura 29 - Gráficos de dispersão, médias e ajuste do modelo Chanter aos dados de (a) comprimento e (b) diâmetro do fruto do cacaueiro ao longo do tempo	63
Figura 30 - Médias, ajuste do modelo Chanter aos dados de (a) comprimento e (b) diâmetro do fruto do cacaueiro ao longo do tempo e ponto de inflexão . .	64

LISTA DE TABELAS

Tabela 1 - Cenário I: Valor médio da porcentagem de viés, viés médio e erro quadrático médio dos estimadores dos parâmetros β_1 , β_2 , β_3 e β_4 do modelo Chanter para os dados simulados de comprimento e diâmetro de frutos de cacaueiro	49
Tabela 2 - Cenário 2: Valor médio da porcentagem de viés, viés médio e erro quadrático médio dos estimadores dos parâmetros β_1 , β_2 , β_3 e β_4 do modelo Chanter para os dados simulados de comprimento e diâmetro de frutos de cacaueiro	55
Tabela 3 - Estimativas dos parâmetros (erros-padrão) e p-valor do teste de Shapiro-Wilk dos modelos Chanter, Logístico e Gompertz aos dados de comprimento de frutos de cacaueiro	58
Tabela 4 - Estimativas dos parâmetros (erros-padrão) e p-valor do teste de Shapiro-Wilk dos modelos Chanter, Logístico e Gompertz aos dados de diâmetro de frutos de cacaueiro	58
Tabela 5 - Avaliadores da qualidade de ajuste dos modelos Chanter, Logístico e Gompertz aos dados de comprimento e diâmetro de frutos de cacaueiro	60
Tabela 6 - Diferença entre os valores de AICc (valor do Peso de Akaike) dos modelos Chanter (C), Logístico (L) e Gompertz (G) aos dados de comprimento e diâmetro de frutos de cacaueiro	61
Tabela 7 - Valores da medida de não linearidade curvatura de Bates e Watts para os modelos Chanter, Logístico e Gompertz aplicados aos dados de comprimento e diâmetro de frutos de cacaueiro	62
Tabela 8 - Valores em porcentagem da medida de não linearidade vício de Box para os modelos Chanter, Logístico e Gompertz aos dados de comprimento e diâmetro de frutos de cacaueiro	62

1 INTRODUÇÃO

Os modelos clássicos de crescimento, como os modelos Logístico (Nelder, 1961) e Gompertz (Laird *et al.*, 1965), têm sido muito utilizados para descrever vários processos biológicos animais e vegetais. Eles podem ser ajustados a dados com relativa facilidade, usando *software* estatístico padrão, e seus parâmetros exibem considerável interpretabilidade biológica.

Uma característica comum entre os modelos clássicos de crescimento é que o crescimento do indivíduo é sempre crescente e tende a um valor assintótico. Geralmente é razoável admitir que, à medida que o tempo (ou outra variável independente) aumenta, o valor da variável dependente aproxima-se de uma constante que não é zero.

As equações dos modelos de crescimento são soluções de equações diferenciais. No caso do modelo Chanter sua equação diferencial apresenta parte da equação diferencial do modelo Logístico e parte da equação diferencial do modelo Gompertz. Esta particularidade indica que o modelo Chanter é um híbrido dos modelos Logístico e Gompertz, cujos parâmetros podem ser interpretados similarmente (FRANCE; THORNLEY, 1984).

O objetivo geral deste trabalho consiste em estudar o modelo Chanter quanto aos seus parâmetros. Realizar um estudo de simulação com o intuito de comparar os modelos Chanter, Logístico e Gompertz, ajustar o modelo Chanter, assim como o Logístico e o Gompertz, a um conjunto de dados do fruto do cacaueteiro (BRITO; SILVA, 1983), fazer a análise de resíduos e diagnósticos são os objetivos específicos.

Para comparar o ajuste destes três modelos serão utilizadas como medidas estatísticas para qualidade do ajuste o critério de informação de Akaike (AIC), o critério de informação de Bayes (BIC), o desvio padrão residual (DPR), as medidas de não linearidade (vício de Box e curvatura de Bates e Watts) e o critério peso de Akaike

Este trabalho está estruturado da seguinte forma: introdução, revisão bibliográfica, material e métodos, resultados e discussão e conclusão.

2 REVISÃO BIBLIOGRÁFICA

2.1 Regressão

Sir Francis Galton em 1889 utilizou pela primeira vez o termo regressão, quando estudou a relação existente entre a estatura dos pais e a estatura dos filhos e notou que os filhos de pais altos tendiam a ser altos, mas não tanto quanto seus pais, e que filhos de pais de menor estatura tendiam a ser baixos, mas novamente não tanto quanto seus pais. Na regressão linear, o termo linear indica que o modelo é linear nos parâmetros.

O modelo de regressão linear pode ser expresso como

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \epsilon, \quad (1)$$

em que Y é a variável resposta, $X_1, X_2, \dots, X_k$ são as variáveis explicativas e $\beta_0, \beta_1, \dots, \beta_k$ são os parâmetros do modelo, que quantificam a influência que uma variável explicativa tem sobre a variável resposta. Para estimar os β 's em (1) usaremos uma amostra de n observações de Y associados com os X 's. O modelo para a i -ésima observação é dado por:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \epsilon_i, \quad (2)$$

em que $i = 1, 2, \dots, n$.

O erro ϵ_i possui as seguintes suposições:

1. $E(\epsilon_i) = 0$ para todo $i = 1, 2, \dots, n$.
2. $\text{Var}(\epsilon_i) = \sigma^2$ para todo $i = 1, 2, \dots, n$.
3. $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ para todo $i \neq j$.

Escrevendo (2) considerando conjuntamente as n observações temos o modelo na forma matricial como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (3)$$

em que $\mathbf{y}$ e $\boldsymbol{\epsilon}$ são vetores $n \times 1$, $\boldsymbol{\beta}$ é um vetor de parâmetros, $p \times 1$ em que $p = k + 1$ e a matriz $\mathbf{X}$ é $n \times p$, assumindo que $n > p$.

As suposições anteriores sobre ϵ_i podem ser expressas em termos do modelo (3) como:

1. $E(\boldsymbol{\epsilon}) = \mathbf{0}$.
2. $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$.¹

Existem vários métodos para a estimação dos parâmetros de um modelo de regressão, o mais usual é o método dos mínimos quadrados desenvolvido por Carl Friedrich Gauss em 1809, o qual consiste na minimização da soma dos quadrados dos resíduos ou da distância entre os valores observados Y_i e os valores preditos $\hat{Y}_i$ pelo modelo ajustado.

A solução de mínimos quadrados consiste nos valores de $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ (chamados de estimadores de mínimos quadrados) para os quais a soma de quadrado dos resíduos (SQR) torna-se mínima, isto é, devemos calcular a diferencial de $\sum_{i=1}^n \epsilon_i^2$ em relação a cada β_j e igualar os resultados a zero para produzir $(k + 1)$ equações que podem ser resolvidas simultaneamente para os β_j 's.

Considerando o modelo dado por:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \epsilon_i \Rightarrow$$

$$\epsilon_i = Y_i - (\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}) = Y_i - E(Y_i), \quad (4)$$

temos que a soma de quadrados dos resíduos é dada por:

$$SQR = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - E(Y_i))^2 = \sum_{i=1}^n [Y_i - (\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})]^2. \quad (5)$$

Considerando $\boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, podemos ainda escrever (5) na forma matricial, como se segue:

¹A suposição $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$ inclui as suposições $\text{Var}(\epsilon_i) = \sigma^2$ e $\text{Cov}(\epsilon_i, \epsilon_j) = 0$.

$$\begin{aligned}
SQR &= \sum_{i=1}^n \epsilon_i^2 = \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}.
\end{aligned}$$

Assim, a diferencial de $\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}$ em relação a $\boldsymbol{\beta}$ é dada por:

$$\frac{\partial \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}}{\partial \boldsymbol{\beta}} = \mathbf{0} - 2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}. \quad (6)$$

Igualando o resultado a zero, obtemos o sistema de equações normais:

$$\mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y} \quad (7)$$

Portanto, $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ é o estimador de mínimos quadrados para o modelo de regressão linear. O termo linear é utilizado para indicar que o modelo é linear em relação aos seus parâmetros $\beta_0, \beta_1, \dots, \beta_k$, ou seja, as derivadas parciais da função em relação aos parâmetros do modelo não dependem destes. Existem situações em que o modelo linear não é o mais apropriado, nestes casos os modelos de regressão não linear podem ser utilizados.

2.2 Regressão não linear

Segundo Smyth (2006) um modelo de regressão não linear pode ser escrito como:

$$Y_i = f(X_i, \boldsymbol{\beta}) + \epsilon_i, i = 1, \dots, n. \quad (8)$$

em que Y_i é a variável resposta; $f(X_i, \boldsymbol{\beta})$ é uma função não linear, com forma conhecida, que depende da variável explicativa x_i e do vetor de parâmetros desconhecidos $\boldsymbol{\beta}$ e ϵ_i é o erro experimental suposto independente e identicamente distribuído, de uma distribuição normal com média 0 e variância σ^2 , isto é, $\epsilon_i \sim N(0; \sigma^2)$.

Matricialmente, o modelo de regressão não linear pode ser expresso como

$$\mathbf{y} = f(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\epsilon}, \quad (9)$$

em que $\mathbf{y}$ é um vetor $(n \times 1)$ de observações de uma variável aleatória Y , $f(\mathbf{X}, \boldsymbol{\beta})$ é um vetor $(n \times 1)$ de funções não lineares em $\boldsymbol{\beta}$ e $\boldsymbol{\epsilon}$ é um vetor $(n \times 1)$ de erros. Dessa forma, considerando o modelo ajustado dado por:

$$\hat{\mathbf{y}} = f(\mathbf{X}, \boldsymbol{\beta}) \quad (10)$$

temos que o vetor de resíduos é dado por $\boldsymbol{\epsilon} = \mathbf{y} - \hat{\mathbf{y}}$. E assim temos que a soma de quadrados dos resíduos, SQR , é dada por $\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}$ pode ser expressa como:

$$SQR = [\mathbf{y} - f(\mathbf{X}, \boldsymbol{\beta})]^\top [\mathbf{y} - f(\mathbf{X}, \boldsymbol{\beta})]. \quad (11)$$

Logo,

$$\begin{aligned} SQR &= \mathbf{y}^\top - [f(\mathbf{X}, \boldsymbol{\beta})]^\top [\mathbf{y} - f(\mathbf{X}, \boldsymbol{\beta})] \\ &= \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top f(\mathbf{X}, \boldsymbol{\beta}) - [f(\mathbf{X}, \boldsymbol{\beta})]^\top \mathbf{y} + [f(\mathbf{X}, \boldsymbol{\beta})]^\top f(\mathbf{X}, \boldsymbol{\beta}) \\ &= \mathbf{y}^\top \mathbf{y} - [f(\mathbf{X}, \boldsymbol{\beta})]^\top \mathbf{y} - [[f(\mathbf{X}, \boldsymbol{\beta})]^\top \mathbf{y} + [f(\mathbf{X}, \boldsymbol{\beta})]^\top f(\mathbf{X}, \boldsymbol{\beta})] \\ &= \mathbf{y}^\top \mathbf{y} - 2[f(\mathbf{X}, \boldsymbol{\beta})]^\top \mathbf{y} + [f(\mathbf{X}, \boldsymbol{\beta})]^\top f(\mathbf{X}, \boldsymbol{\beta}). \end{aligned}$$

Assim como nos modelos de regressão linear, uma das formas de estimar os parâmetros dos modelos não lineares consiste em fazer mínima a soma dos quadrados residuais, conhecida como método dos mínimos quadrados. Para tanto é necessário fazer a diferenciação de SQR com relação a cada um dos parâmetros, isto é, $\frac{\partial \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}}{\partial \boldsymbol{\beta}}$ e igualar as derivadas a zero obtendo p equações normais não lineares. Geralmente o sistema de equações normais não lineares não possui solução analítica, assim, a obtenção das estimativas dos parâmetros é realizada utilizando métodos numéricos iterativos. Na literatura são apresentados vários métodos numéricos iterativos. São exemplos, método de Newton-Raphson,

algoritmo EM (DEMPSTER ET AL., 1977), método de Gauss-Newton, entre outros.

2.3 Método iterativo de Gauss-Newton

Segundo Mazucheli e Achcar (2002), o algoritmo de estimação dos parâmetros em regressão não linear mais comum e que será utilizado neste trabalho, é o método de Gauss-Newton, que se baseia em uma aproximação linear, de primeira ordem, em série de Taylor para a função $f(\mathbf{X}, \boldsymbol{\beta}) = (f(x_1, \boldsymbol{\beta}), \dots, f(x_n, \boldsymbol{\beta}))^\top$ em torno de uma vizinhança de $\boldsymbol{\beta}^{(a)}$, em que $\boldsymbol{\beta}^{(a)}$ é o verdadeiro valor do parâmetro $\boldsymbol{\beta}$. Essa aproximação linear é dada por:

$$f(x_i, \boldsymbol{\beta}) \approx f(x_i, \boldsymbol{\beta}^{(a)}) + \sum_{j=1}^p \left[\frac{\partial f(x_i, \boldsymbol{\beta})}{\partial \beta_j} \right]_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(a)}} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(a)}). \quad (12)$$

Em notação matricial temos:

$$\mathbf{f}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}^{(a)}) + \frac{\partial \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}^{(a)})}{\partial \boldsymbol{\beta}_j^{(a)\top}} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(a)}) \quad (13)$$

Usando (13), pode-se escrever a soma de quadrado dos resíduos, definida em (11), na forma:

$$SQR = [\mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta})]^\top [\mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta})] \quad (14)$$

$$= \|\mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta})\|^2 \quad (15)$$

$$\approx \|\mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}^{(a)}) - \frac{\partial \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}^{(a)})}{\partial \boldsymbol{\beta}_j^{(a)\top}} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(a)})\|^2. \quad (16)$$

Fazendo $\mathbf{z} = \mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}^{(a)})$ e $\boldsymbol{\beta}^* = \boldsymbol{\beta} - \boldsymbol{\beta}^{(a)}$. Tem-se que

$$SQR = \left\| \mathbf{z} - \frac{\partial \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}^{(a)})}{\partial \boldsymbol{\beta}_j^{(a)\top}} \boldsymbol{\beta}^* \right\|^2.$$

Logo (14) será minimizada quando β^* for dada por:

$$\widehat{\beta^*} = \left[\left(\frac{\partial \mathbf{f}(\mathbf{X}, \beta^{(a)})}{\partial \beta_j^{(a)\top}} \right)^\top \left(\frac{\partial \mathbf{f}(\mathbf{X}, \beta^{(a)})}{\partial \beta_j^{(a)\top}} \right) \right]^{-1} \left(\frac{\partial \mathbf{f}(\mathbf{X}, \beta^{(a)})}{\partial \beta_j^{(a)\top}} \right)^\top \mathbf{z}. \quad (17)$$

A cada iteração é gerada uma nova aproximação $\beta^{(a+1)}$ para o vetor de parâmetros $\beta^{(a)}$, através de

$$\beta^{(a+1)} = \beta^{(a)} + \left[\left(\frac{\partial \mathbf{f}(\mathbf{X}, \beta^{(a)})}{\partial \beta_j^{(a)\top}} \right)^\top \left(\frac{\partial \mathbf{f}(\mathbf{X}, \beta^{(a)})}{\partial \beta_j^{(a)\top}} \right) \right]^{-1} \left(\frac{\partial \mathbf{f}(\mathbf{X}, \beta^{(a)})}{\partial \beta_j^{(a)\top}} \right)^\top \mathbf{z} \quad (18)$$

O processo apresentado na equação (18) é repetido até que se obtenha convergência definida por algum critério, como $(\beta^{(a+1)} - \beta^{(a)}) < 0,000001$, por exemplo.

O método de Gauss-Newton equivale a uma simplificação do método de Newton-Raphson, necessitando de um número maior de iterações que este, entretanto os cálculos realizados em cada iteração são mais simples.

2.4 Modelos de crescimento

Segundo Seber e Wild (2003), a análise dos dados sobre crescimento é importante em muitas áreas de pesquisa. Na biologia, o interesse está na descrição do crescimento de animais e plantas na tentativa de compreender seus mecanismos essenciais. Os químicos estão interessados na formulação do produto de uma reação química ao longo do tempo. Na agricultura, são óbvias as vantagens econômicas e administrativas em conhecer como a produção cresce, o quão rápido ela cresce, e como esse fator, crescimento, responde a condições ambientais ou tratamentos.

Há vários estudos utilizando medidas de comprimento e diâmetro de frutos, como Alexandre et al. (2015), o qual estudou o cacau e concluiu que as variáveis referentes ao tamanho do fruto, comprimento e diâmetro foram as que mais contribuíram na diversidade genética dos genótipos.

Dentre os inúmeros modelos usados nesses estudos destacam-se os modelos de regressão não linear utilizados para descrever curvas de crescimento: Logístico, Gompertz

e Chanter. Esses três modelos de regressão não linear são exemplos de curvas sigmóides, cujos gráficos têm a forma de S.

2.4.1 Modelo Logístico

O modelo Logístico foi proposto por Pierre François Verhulst em 1838, aplicado para descrever curvas de crescimento individual, populacional (PINHEIRO; BATES, 2000; SCHABENBERGER; PIERCE, 2002) e o progresso de doenças (SEGARRA et al., 2001).

Uma aplicação típica da função Logística é um modelo comum do crescimento populacional, onde a taxa de reprodução é proporcional tanto à população existente quanto à quantidade de recursos disponíveis, mantendo-se constantes todos os outros aspectos relacionados ao crescimento. A população cresce até um limite sustentável, isto é, ela tende a se estabilizar (TAVONI; OLIVEIRA, 2013).

Considere a seguinte parametrização do modelo Logístico:

$$y(x) = \frac{\beta_1}{1 + e^{-\beta_2 - \beta_3 x}} + \epsilon. \quad (19)$$

em que β_1, β_2 e β_3 são os parâmetros deste modelo.

Nessa parametrização do modelo Logístico os parâmetros β_1 e β_3 são positivos e o parâmetro β_2 pode assumir valores positivos ou valores negativos. A equação do modelo Logístico é solução da seguinte equação diferencial, cuja demonstração encontra-se no apêndice deste trabalho:

$$\frac{dy(x)}{dx} = \beta_2 y(x) \left(1 - \frac{y(x)}{\beta_1} \right). \quad (20)$$

A partir de agora trataremos o modelo Logístico de função Logística, sendo assim, a equação diferencial corresponde a dizer que a taxa relativa de crescimento $\left(\frac{dy(x)}{dx} \right)$, é proporcional ao valor presente (y) e à quantidade que falta para a função atingir o parâmetro β_1 , parâmetro associado à assíntota. Segundo Michan e Pinho (2014), fazendo-se um estudo da função Logística, obtém-se uma interpretação prática dos parâmetros. Observe que $y = \beta_1$ é uma assíntota horizontal à direita, e $y = 0$ é uma assíntota horizontal à esquerda. A função $y(x)$ é estritamente crescente cujo ponto de inflexão é $\left(-\frac{\beta_2}{\beta_3}, \frac{\beta_1}{2} \right)$. Segundo Guidorizzi (1997), dizemos que $(p, f(p))$ é um ponto de inflexão do gráfico da

função f se existirem números reais a e b , com $p \in (a, b) \subset D_f$, tal que f tenha concavidades contrárias em (a, p) e em (p, b) . No caso da função Logística, o gráfico da função tem concavidade para cima se $x < -\frac{\beta_2}{\beta_3}$ e concavidade para baixo se $x > -\frac{\beta_2}{\beta_3}$. Fazendo-se $x = 0$ obtém-se $\beta_2 = \ln\left(\frac{y_0}{\beta_1 - y_0}\right)$, em que y_0 é o valor inicial de y . A função $y(x)$ também caracteriza-se por ser simétrica em relação ao seu ponto de inflexão. Os gráficos da função Logística, para alguns valores arbitrários dos parâmetros podem ser visualizados nas Figuras (1), (2) e (3).

Figura 1 - Gráficos da função Logística para diferentes valores dos parâmetros β_1 sendo (a) $\beta_2 = -1$ e $\beta_3 = 2$ e (b) $\beta_2 = 1$ e $\beta_3 = 2$

Figura 2 - Gráficos da função Logística para diferentes valores dos parâmetros β_2 sendo $\beta_1 = 2$ e $\beta_3 = 2$

Figura 3 - Gráficos da função Logística para diferentes valores dos parâmetros β_3 sendo (a) $\beta_1 = 3$ e $\beta_2 = -1$ e (b) $\beta_1 = 3$ e $\beta_2 = 1$

Observe na Figura (1) que independentemente ao valor de β_2 , positivo ou

negativo, o parâmetro β_1 é uma assíntota superior à direita. Porém, no caso (a), em que β_2 é negativo, o intercepto é menor do que no caso (b) em que β_2 é positivo. Na Figura (2), note que variando o β_2 temos variação do intercepto.

Observe na Figura (3) que independentemente do valor de β_2 , positivo ou negativo, ao diminuir o parâmetro β_3 , a curva atinge as assíntotas mais lentamente. Porém, no caso (a), em que β_2 é negativo, o intercepto é menor do que no caso (b) em que β_2 é positivo.

Quando o fruto atinge metade do seu tamanho máximo assintótico (β_1), a velocidade de crescimento atinge seu valor máximo decrescendo daí em diante tendendo para zero. A relação $\left(-\frac{\beta_2}{\beta_3}\right) = x$ indica que em x dias ocorreu a velocidade máxima do crescimento por dia, após esse dia há uma desaceleração no crescimento do fruto.

Segundo Michan e Pinho (2014), interpreta-se os parâmetros como sendo, que: β_1 é a capacidade de sustentação do meio, β_2 é a taxa relativa máxima de crescimento, correspondente à fase de arranque do crescimento e β_3 é a razão entre a quantidade que falta para atingir a assíntota β_1 e o valor inicial y_0 .

2.4.2 Modelo Gompertz

O modelo sugerido por Benjamin Gompertz em 1825 teve aplicação inicial na descrição em tabelas atuariais, que contém as probabilidades de sobrevivência e morte de uma população. Essas tabelas são utilizadas para o cálculo de benefícios de renda e seguro de vida (Renda Vitalícia), mas também é muito aplicado em estudos de crescimento de animais (FREITAS, 2005), plantas (MENDES, 2011) e tumores (DOMINGUES, 2011). O modelo Gompertz é bastante semelhante ao modelo Logístico, possui um crescimento inicial exponencial e, a seguir, assintótico (MISCHAN; PINHO, 2014). A equação do modelo Gompertz é dada por

$$y(x) = \beta_1 e^{(-e^{(-\beta_2(x-\beta_3))})} + \epsilon \quad (21)$$

em que β_1, β_2 e β_3 são parâmetros positivos.

A equação do modelo Gompertz é solução da seguinte equação diferencial e cuja solução também encontra-se no apêndice deste trabalho:

$$\frac{dy(x)}{dx} = \beta_2 y(x) e^{-\beta_2(x-\beta_3)} \quad (22)$$

A equação diferencial corresponde a dizer que a taxa relativa de crescimento $\left(\frac{dy(x)}{dx}\right)$ é proporcional a y e à uma função exponencial negativa que depende de x . Semelhante ao Logístico, no entanto se distingue do Logístico por descrever uma taxa de crescimento cuja taxa relativa decresce exponencialmente como função do tempo. Fazendo um estudo da equação do modelo Gompertz, observa-se que $y = \beta_1$ é uma assíntota horizontal à direita, e $y = 0$ é uma assíntota horizontal à esquerda. A função $y(x)$ é estritamente crescente, tem concavidade para cima se $x < \beta_3$, concavidade para baixo se $x > \beta_3$ e seu ponto de inflexão é $(\beta_3, \frac{\beta_1}{e})$. Fazendo-se $x = 0$ obtém-se $\beta_2 = \frac{\ln[\ln(\frac{\beta_1}{y_0})]}{\beta_3}$, em que y_0 é o valor inicial de y . Os gráficos da função Gompertz, para alguns valores arbitrários de seus parâmetros pode ser visualizado na Figura (4).

Os modelos Logístico e Gompertz possuem semelhanças e diferenças quanto aos seus parâmetros e interpretações práticas. O parâmetro β_1 em ambos os modelos representa uma assíntota superior à direita. O modelo Logístico é simétrico e o modelo Gompertz é assimétrico em relação aos seus respectivos pontos de inflexão. Esta assimetria indica que no modelo Logístico o fruto tem uma desaceleração no crescimento quando atinge metade do seu tamanho máximo assintótico $\frac{\beta_1}{2}$, já no modelo Gompertz o fruto tem uma desaceleração no crescimento quando atinge $\frac{\beta_1}{e}$, ou seja, no modelo Gompertz essa mudança de aceleração no crescimento acontece quando o fruto atinge aproximadamente 37% de seu valor assintótico. Essa informação é interessante para descobrir a melhor época de colheita, há certas variedades que necessitam ser colhidas mais tardiamente que outras.

Figura 4 - Representações gráficas da função Gompertz sendo (a) $\beta_2 = 1$ e $\beta_3 = 2$ fixos e β_1 variando, (b) $\beta_1 = 4$ e $\beta_3 = 2$ fixos e β_2 variando e (c) $\beta_1 = 3$ e $\beta_2 = 1$ fixos e β_3 variando

Observe em (a) que o parâmetro β_1 representa a assíntota superior à direita, em (b), note que quanto maior o valor do parâmetro β_2 temos que a função atinge a assíntota

mais lentamente e no caso (c) o parâmetro β_3 representa o instante em que a função muda de concavidade.

Segundo Michan e Pinho (2014), interpreta-se os parâmetros como sendo, que: β_1 é a capacidade de sustentação do meio, β_2 fornece a ideia da razão entre a assíntota β_1 e a quantidade que falta para atingi-la e β_3 é a abscissa do ponto de inflexão .

2.4.3 Modelo Chanter

O modelo Chanter foi proposto por Dennis Osborne Chanter em 1976. É um híbrido entre os modelos Logístico e o Gompertz cujos parâmetros podem ser interpretados similarmente (FRANCE; THORNLEY, 1984). O gráfico da função Chanter, para alguns valores arbitrários dos parâmetros pode ser visualizado na Figura (5). O estudo da função para diferentes valores dos parâmetros será apresentado na seção material e métodos.

Figura 5 - Representação gráfica da função Chanter para $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 0,5$ e $\beta_4 = -1$

O modelo Chanter é adequado à representação de dados de crescimento de organismos vivos, onde o crescimento inicial é exponencial e, a seguir, assintótico. A função Chanter pode ser representada por

$$y(x) = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1) e^{(-\frac{\beta_3}{\beta_4}(1 - e^{-\beta_4 x}))}}, \quad (23)$$

em que $\beta_1, \beta_2, \beta_3$ e β_4 são os parâmetros do modelo.

A equação que representa o modelo Chanter é solução da seguinte equação diferencial

$$\frac{dy(x)}{dx} = \beta_3 y(x) \left(1 - \frac{y(x)}{\beta_2} \right) e^{-\beta_4 x} \quad (24)$$

A equação diferencial corresponde a dizer que a taxa relativa de crescimento $\left(\frac{dy(x)}{dx} \right)$, é proporcional a y , à quantidade que falta para a função atingir o parâmetro β_2 (comum ao modelo Logístico) e à uma função exponencial negativa que depende de x (comum ao modelo Gompertz). Por esse motivo é considerado um híbrido entre o modelo Logístico e o modelo Gompertz.

2.5 Critérios para qualidade do ajuste

Para comparar os três modelos, Logístico, Gompertz e Chanter, serão utilizadas como medidas estatísticas para qualidade do ajuste, o Critério de informação de Akaike (AIC), o Peso de Akaike, o critério de informação de Bayes (BIC), o desvio padrão residual (DPR) e as medidas de não linearidade (vício de Box e curvatura de Bates e Watts).

Critério de informação de Akaike(AIC)

O critério de informação de Akaike avalia a qualidade do ajuste entre dois ou mais modelos. Akaike (1974) definiu seu critério de informação como:

$$AIC = -2\ln(L) + 2p \quad (25)$$

sendo L o máximo da função de verossimilhança que sob normalidade coincide com o mínimo da soma de quadrado dos resíduos e p o número de parâmetros ajustados. O modelo com menor valor de AIC é considerado o modelo de melhor ajuste.

Uma maneira de quantificar a chance de que o modelo esteja correto, é através de probabilidades chamadas Pesos de Akaike (MOTULSKY; CHRISTOPOULOS, 2003). Para calcular essas probabilidades é necessário ter o valor de AIC corrigido. Bozdogan (1987), propôs a seguinte correção para o AIC:

$$AICc = -2\log(L) + 2p + 2\frac{p(p+1)}{n-p-1}. \quad (26)$$

Se os valores de AICc de dois modelos concorrentes são muito próximos, não há muita evidência para escolher um modelo ou outro. Se os valores de AICc são distantes, então a evidência é esmagadora.

Peso de Akaike

A porcentagem de chance de um modelo ser correto é calculado pela seguinte equação, onde Δ é a diferença entre os valores de AICc.

$$\text{Peso de Akaike} = \frac{e^{-0,5\Delta}}{1 + e^{-0,5\Delta}} \quad (27)$$

A porcentagem de chance de cada modelo ser o correto está representada graficamente na Figura 6.

Figura 6 - Representação gráfica da relação entre a diferença nos valores de AICc e a probabilidade de que cada modelo seja verdadeiro (Fonte: MOTULSKY; CHRISTOPOULOS, 2003, p.145)

O peso de Akaike calcula a probabilidade relativa de dois modelos e pode ser estendido para uma família de três ou mais modelos. Vale lembrar que é sempre possível que outro modelo seja mais provável e este pode não ter sido escolhido.

Critério de informação de Bayes (BIC)

O Critério de Informação Bayesiano (BIC), proposto por Schwarz (1978) é utilizado para maximizar a probabilidade de escolha do modelo verdadeiro. O critério é

dado por:

$$BIC = -2\ln(L) + p\ln(n) \quad (28)$$

sendo L o máximo da função de verossimilhança e p o número de parâmetros ajustados. O modelo com menor valor de BIC é considerado o de melhor ajuste. O critério BIC penaliza mais fortemente modelos com um maior número de parâmetros do que o AIC tendendo, dessa forma, a selecionar modelos com um menor número de parâmetros.

Desvio padrão residual (DPR)

O desvio padrão residual (DPR) quantifica o tamanho médio dos resíduos, alguns livros referem-se ao DPR como “*root mean square*” ou raiz do quadrado médio (MOTULSKY; CHRISTOPOULOS, 2003). O DPR é calculado a partir do quadrado médio do resíduo (QMR), do número de observações n e do número de parâmetros p do modelo, usando a seguinte equação:

$$DPR = \sqrt{\frac{QMR}{n - p}} \quad (29)$$

Vício de Box e curvatura de Bates e Watts

As duas medidas, vício de Box e curvatura de Bates e Watts, são empregadas na avaliação da não linearidade de um modelo, podendo também ser usadas em problemas de discriminação, uma vez que o melhor modelo possível, dentre todos os propostos, seria aquele que apresenta um comportamento mais próximo do linear (RATKOWSKY, 1983). Um modelo deve ser preferido em relação a outro se apresentar os menores valores das medidas de não linearidade (ZEVIANI, 2012).

As medidas de não linearidade em regressão não linear estão relacionadas ao método de minimização dos erros, que utiliza uma expansão em série de Taylor. Souza (1998) cita que quanto mais linearizável for $f(\mathbf{X}, \hat{\beta})$ melhores são os resultados inferenciais associados ao modelo não linear. As duas medidas de não linearidade mais utilizadas são: o vício de Box (Box, 1971) e medidas de curvatura de Bates e Watts (Bates & Watts, 1980). Ambas as medidas devem assumir os menores valores possíveis para que a função

seja considerada linearizável.

O vício de Box

Seja $\mathbf{f}_i(x, \boldsymbol{\beta}), i = 1 \dots n$ um vetor de funções não lineares, $n \times 1$ e $\boldsymbol{\beta}$ um vetor de parâmetros, $p \times 1$. Segundo Mazucheli e Achcar (2002) seja $F_i(\boldsymbol{\beta})$ com $i = 1, \dots, n$ o vetor das derivadas primeiras de ordem $p \times 1$ e $H_i(\boldsymbol{\beta})$ a matriz das derivadas segundas de ordem $p \times p$, de $\mathbf{f}_i(x, \boldsymbol{\beta})$ ambas com respeito aos elementos de $\boldsymbol{\beta}$. O vício de Box é dado por:

$$V(\hat{\boldsymbol{\beta}}) = \frac{-\sigma^2}{2} \left[\sum_{i=1}^n F_i(\boldsymbol{\beta}) F_i^\top(\boldsymbol{\beta}) \right]^{-1} \sum_{i=1}^n F_i(\boldsymbol{\beta}) \text{tr} \left[\left(\sum_{i=1}^n F_i(\boldsymbol{\beta}) F_i^\top(\boldsymbol{\beta}) \right)^{-1} H_i(\boldsymbol{\beta}) \right]. \quad (30)$$

Segundo Mazucheli e Achcar (2002) para o cálculo de 30 usa-se $\hat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2$ como sendo os verdadeiros valores de $\boldsymbol{\beta}$ e σ^2 respectivamente. Assim, o vetor $p \times 1$ determinado por 30 representa a discrepância entre as estimativas dos parâmetros e seus verdadeiros valores.

A porcentagem do vício é obtida por:

$$\%V(\hat{\boldsymbol{\beta}}) = \frac{100V(\hat{\boldsymbol{\beta}})}{\hat{\boldsymbol{\beta}}} \quad (31)$$

Segundo Box (1971), quando os valores de vício são superiores a 1%, considera-se que o modelo tem comportamento não linear. Percebe-se, portanto, que quanto menor for a diferença entre a aproximação linear e a quadrática, menor será $\%V(\hat{\boldsymbol{\beta}})$.

Curvatura de Bates e Watts

Bates e Watts (1980) decompuseram uma medida de curvatura em duas partes: uma componente tangencial (K^T) que é paramétrica e uma componente normal (K^N) que é intrínseca. Essas medidas de curvatura são calculadas através das seguintes expressões:

$$K^T = \frac{\|(\beta - \hat{\beta})^T H^T(\beta - \hat{\beta})\|}{\|F(\hat{\beta})(\beta - \hat{\beta})\|^2} \quad e \quad K^N = \frac{\|(\beta - \hat{\beta})^T H^N(\beta - \hat{\beta})\|}{\|F(\hat{\beta})(\beta - \hat{\beta})\|^2} \quad (32)$$

em que H^T indica a componente tangencial e H^N indica a componente normal. Essa decomposição pode ser obtida definindo-se uma matriz de projeção $P_{F(\hat{\beta})}$, dada por:

$$F(\beta)(F(\beta)'F(\beta))^{-1}F(\beta)'.$$

Assim, K^T é a curvatura paramétrica e K^N é a curvatura intrínseca. Vale lembrar que T é uma notação para componente tangencial e não para transposto.

O interessante é que estes valores de K^T e K^N sejam pequenos, indicando linearidade do modelo. Estes valores podem ser comparados com a estatística (SOUZA, 1998):

$$\frac{1}{2\sqrt{F_{1-\alpha}(p, (n-p))}}. \quad (33)$$

onde $F_{1-\alpha}(p, (n-p))$ é o valor crítico obtido a partir de uma distribuição F com p e $(n-p)$ graus de liberdade.

A abordagem de Bates e Watts (1980) quantifica a não linearidade do modelo, baseando-se no local da solução, enquanto a abordagem de Box (1971) fez uso de medidas associadas aos parâmetros estimados.

3 MATERIAL E MÉTODOS

3.1 Material

3.1.1 Conjunto de dados

Os dados utilizados no presente trabalho são referentes ao comprimento (cm) e diâmetro (cm) dos frutos do cacauzeiro (BRITO; SILVA, 1983). Os frutos procederam de cacauzeiros sombreados do clone SIAL - 105, que estavam dispostos em fileiras no Jardim Clonal da Quadra E, do Centro de Pesquisa do Cacau, em Ilhéus, Bahia. As 23 árvores foram polinizadas artificialmente, tomando-se pólen do clone EEG-9 que estavam na fileira vizinha, o terreno se caracterizava por ser do tipo alfisol, solos com maior conteúdo de argila, mas sem húmus, da série germoplasma.

A primeira coleta dos frutos iniciou-se um mês após a polinização e repetiu-se a cada 15 dias. Como a princípio os frutos eram pequenos, na primeira coleta foram tomados 50 frutos. Posteriormente, com o aumento do tamanho dos frutos e por escassez de material, a quantidade de frutos por coleta foi diminuindo, tomaram-se 40 frutos na segunda coleta, 30 frutos, na terceira e na quarta, 20, a quinta, sexta e oitava coleta, 15 frutos, na nona e décima e por fim 16 na décima primeira. Após a coleta, os frutos eram colocados em sacos plásticos e imediatamente levados ao laboratório, onde eram tomadas as medidas de crescimento utilizando um paquímetro.

3.1.2 Modelos

A esses dados foram ajustados os modelos não lineares Logístico, Gompertz e Chanter, com as seguintes parametrizações:

$$\text{Modelo Logístico: } y_i = \frac{\beta_1}{1 + e^{-\beta_2 - \beta_3 x_i}} + \epsilon_i.$$

$$\text{Modelo Gompertz: } y_i = \beta_1 e^{(-e^{(-\beta_2(x_i - \beta_3))})} + \epsilon_i$$

$$\text{Modelo Chanter: } y_i = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1) e^{(-\frac{\beta_3}{\beta_4}(1 - e^{-\beta_4 x_i}))}} + \epsilon_i$$

em que: $i = 1, 2, \dots, n$; $\beta_1, \beta_2, \beta_3$ e β_4 são parâmetros dos modelos e ϵ_i é o erro aleatório suposto independente e identicamente distribuído seguindo uma distribuição normal com

média 0 e variância σ^2 .

A obtenção das estimativas dos parâmetros foi realizada por meio do método dos mínimos quadrados, que conforme dito anteriormente, para os modelos de regressão não lineares, o sistema de equações normais não é resolvido facilmente, sendo necessários métodos iterativos na obtenção dessas estimativas. Neste trabalho foi utilizado o Método Iterativo de Gauss-Newton. Os valores iniciais para o ajuste dos modelos Logístico e Gompertz foram escolhidos através da linearização dos próprios modelos, cujo *script* encontra-se no anexo deste trabalho.

3.2 Métodos

3.2.1 Simulação

Para avaliar os estimadores dos parâmetros β_1 , β_2 , β_3 e β_4 do modelo Chanter foi realizado um estudo de simulação que foi dividido em dois cenários, no cenário 1 considerou-se uma amostra aleatória de tamanho 11, esse tamanho foi tomado para que pudesse representar os dados reais estudados, pois o comprimento e o diâmetro dos cacaos foram medidos em 11 ocasiões num período de 180 dias, como apresentado anteriormente.

No cenário 2 considerou-se uma amostra aleatória de tamanho 151 representando a medição diária dos frutos ao longo de 180 dias, iniciando no 30º dia após a polinização da árvore, em ambos os cenários foram consideradas 100 repetições.

Os dados foram simulados por meio do modelo Chanter cujos valores paramétricos β^* foram definidos como valores próximos as estimativas dos parâmetros do modelo Chanter ajustado aos dados reais.

Nesse estudo analisou-se a porcentagem da média do viés e o erro quadrático médio dos estimadores dos parâmetros do modelo Chanter dados pelas seguintes expressões:

(i) Viés (%)

$$V(\beta, \beta^*) = \frac{\beta - \beta^*}{\beta^*} \cdot 100 \quad (34)$$

em que β é a média amostral das estimativas dos parâmetros e β^* é o valor paramétrico. Para verificar se o viés médio é nulo realizou-se um teste de hipóteses bilateral

através da seguinte estatística:

$$Z_{calc} = \frac{\beta - \beta^*}{\sqrt{\frac{s^2}{N}}}, \quad (35)$$

em que β é a média amostral das estimativas dos parâmetros, β^* é o valor paramétrico, s é o desvio padrão residual e N é o tamanho populacional, considerando $Z \sim N(0, 1)$ e $\alpha=0,05$ o nível nominal de significância do teste.

(ii) Erro quadrático médio

$$EQM(\beta, \beta^*) = E[\beta - \beta^*]^2 \quad (36)$$

Facilmente, podemos ver que

$$EQM(\beta, \beta^*) = \text{Var}[\beta] + (V(\beta, \beta^*))^2. \quad (37)$$

De fato,

$$\begin{aligned} E[\beta - \beta^*]^2 &= E[\beta^2 - 2\beta^*\beta + \beta^{*2}] \\ &= E[\beta^2] + E[-2\beta^*\beta] + E[\beta^{*2}] \\ &= E[\beta^2] - 2\beta^*E[\beta] + \beta^{*2} \\ &= E[\beta^2] - (E[\beta])^2 + (E[\beta])^2 - 2\beta^*E[\beta] + \beta^{*2} \\ &= \text{Var}[\beta] + (E[\beta] - \beta^*)^2 \\ &= \text{Var}[\beta] + (V(\beta, \beta^*))^2. \end{aligned}$$

3.2.2 Ajuste empírico dos modelos

As análises foram realizadas com o auxílio do *software* R. Para o ajuste dos modelos Logístico, Gompertz e Chanter foi utilizado a função `nls` do pacote `stats`, cujo *script* encontra-se no anexo deste trabalho. O ajuste foi realizado para as médias dos dados, comprimento e diâmetro, de cada dia de coleta. Na análise dos resíduos foi utilizado o teste de Shapiro-Wilk, e os gráficos de resíduos padronizados e normal de probabilidade

para verificar a normalidade dos resíduos.

Para comparar esses modelos, foram utilizadas como medidas estatísticas para qualidade do ajuste, o Critério de informação de Akaike (AIC), o Peso de Akaike, o critério de informação de Bayes (BIC), o desvio padrão residual (DPR) e as medidas de não linearidade (vício de Box e curvatura de Bates e Watts).

4 RESULTADOS E DISCUSSÃO

4.1 Estudo da função Chanter

Para o modelo Chanter tem-se as seguintes condições de existência:

1. β_1, β_2 e $\beta_3 \in \mathbb{R}_+^*$
2. $\beta_4 \in \mathbb{R}^*$
3. $\beta_2 > \beta_1$

Foi feito um estudo matemático da função Chanter expressa em (23). As assíntotas foram estudadas assim, quando os valores de x crescem ou decrescem infinitamente, temos:

Para $\beta_4 > 0$,

$$\lim_{x \rightarrow \infty} y = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1)e^{-\frac{\beta_3}{\beta_4}}} \quad e \quad \lim_{x \rightarrow -\infty} y = 0. \quad (38)$$

Para $\beta_4 < 0$:

$$\lim_{x \rightarrow \infty} y = \beta_2 \quad e \quad \lim_{x \rightarrow -\infty} y = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1)e^{-\frac{\beta_3}{\beta_4}}}. \quad (39)$$

Portanto, para $\beta_4 > 0$ as retas

$$y = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1)e^{-\frac{\beta_3}{\beta_4}}} \quad e \quad y = 0 \quad (40)$$

são assíntotas horizontais à direita e à esquerda, respectivamente.

E para $\beta_4 < 0$, as retas

$$y = \beta_2 \quad e \quad y = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1)e^{-\frac{\beta_3}{\beta_4}}} \quad (41)$$

são assíntotas horizontais à direita e à esquerda, respectivamente.

Figura 7 - Representações gráficas da função Chanter para $\beta_1 = 2$, $\beta_2 = 3$ e $\beta_3 = 1$, (a) $\beta_4 = 0,5$ e (b) $\beta_4 = -0,5$

Note que na Figura 7 o primeiro gráfico representa a função Chanter com $\beta_4 > 0$ e suas assíntotas são:

$$y = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1) e^{-\frac{\beta_3}{\beta_4}}} = \frac{2 \cdot 3}{2 + (3 - 1) e^{-\frac{1}{0,5}}} = 2,81$$

e

$$y = 0.$$

Já o segundo gráfico representa a função Chanter com $\beta_4 < 0$ e suas assíntotas são:

$$y = \beta_2 = 3$$

e

$$y = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1) e^{-\frac{\beta_3}{\beta_4}}} = \frac{2.3}{2 + (3 - 1) e^{-\frac{-1}{-0.5}}} = 0,64.$$

Continuando o estudo da função Chanter, a derivada primeira da função Chanter é dada pela seguinte expressão:

$$\frac{dy}{dx} = \frac{\beta_1 \beta_2 \beta_3 (\beta_2 - \beta_1) e^{(-\frac{\beta_3}{\beta_4} + \frac{\beta_3}{\beta_4} e^{-\beta_4 x} - \beta_4 x)}}{[\beta_1 + (\beta_2 - \beta_1) e^{(-\frac{\beta_3}{\beta_4} + \frac{\beta_3}{\beta_4} e^{-\beta_4 x})}]^2}. \quad (42)$$

A derivada segunda da função Chanter pode ser escrita da seguinte forma:

$$\frac{d^2 y}{dx^2} = \frac{2\beta_1 \beta_2 \beta_3^2 (\beta_2 - \beta_1)^2 e^{(-2\frac{\beta_3}{\beta_4} + 2\frac{\beta_3}{\beta_4} e^{-\beta_4 x} - 2\beta_4 x)}}{[\beta_1 + (\beta_2 - \beta_1) e^{(-\frac{\beta_3}{\beta_4} + \frac{\beta_3}{\beta_4} e^{-\beta_4 x})}]^3} + \frac{\beta_1 \beta_2 \beta_3 (\beta_2 - \beta_1) e^{(-\frac{\beta_3}{\beta_4} + \frac{\beta_3}{\beta_4} e^{-\beta_4 x} - \beta_4 x)} (-\beta_3 e^{-\beta_4 x} - \beta_4)}{[\beta_1 + (\beta_2 - \beta_1) e^{(-\frac{\beta_3}{\beta_4} + \frac{\beta_3}{\beta_4} e^{-\beta_4 x})}]^2} \quad (43)$$

Para encontrar o ponto de inflexão do modelo Chanter é necessário igualar a derivada segunda a zero, porém, a solução dessa equação não possui forma fechada. Com o auxílio do *software* MAPLE, é possível verificar que a coordenada x do ponto de inflexão não é única, assim a solução da equação fica em função de uma variável global Z e é expressa por:

$$-\frac{\ln\left(\frac{(-Z\beta_4 e^Z \beta_2 + Z\beta_4 e^Z \beta_1 + Z\beta_4 \beta_1 - \beta_3 e^Z \beta_2 + \beta_3 e^Z \beta_1 + \beta_1 \beta_3 + \beta_4 \beta_1 + \beta_4 e^Z \beta_2 - \beta_4 e^Z \beta_1) \beta_4 + \beta_3}{\beta_3}\right)}{\beta_4}. \quad (44)$$

Para $x = 0$, obtemos o valor inicial de y igual a β_1 , ou seja, a função intercepta o eixo y sempre em β_1 como nos gráficos da Figura 8.

Figura 8 - Gráficos da função Chanter para diferentes valores do parâmetro β_1 sendo $\beta_2 = 3$ e $\beta_3 = 1$ (a) $\beta_4 = 0,5$ e (b) $\beta_4 = -0,5$

Na Figura 9 estão três funções de Chanter com valores iguais para β_1 , β_2 e β_4 para o caso com $\beta_4 > 0$ e $\beta_4 < 0$ e diferentes valores de β_3 . Observe que para valores maiores de β_3 a curva é relativamente mais "fechada", com menor valor de β_3 , o crescimento é mais lento, à medida que x aumenta, a curva demora mais para se aproximar da assíntota.

Figura 9 - Gráficos da função Chanter para diferentes valores do parâmetro β_3 sendo $\beta_1 = 2$, $\beta_2 = 3$ (a) $\beta_4 = 0,5$ e (b) $\beta_4 = -0,5$

Na Figura 10 estão três funções de Chanter com os mesmos valores para β_1 , β_2 e β_3 e diferentes valores de β_4 tanto positivos quanto negativos. No caso que β_4 é positivo quanto maior for seu valor, mais rápido a curva atinge suas assíntotas e no caso que β_4 é negativo, quanto menor for seu valor, mais rápido a curva atinge suas assíntotas.

Figura 10 - Gráficos da função Chanter para $\beta_1 = 3$, $\beta_2 = 4$ e $\beta_3 = 2$ e diferentes valores de β_4

Segundo France e Thornley (1984) o modelo Chanter é solução da equação diferencial expressa em (24). De fato, devemos buscar uma solução que satisfaça o seguinte sistema, para o caso de $\beta_4 > 0$:

$$(I) \left\{ \begin{array}{l} \frac{dy(x)}{dx} = \beta_3 y(x) \left(1 - \frac{y(x)}{\beta_2} \right) e^{-\beta_4 x} \\ \lim_{x \rightarrow \infty} y(x) = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1) e^{-\frac{\beta_3}{\beta_4}}} \\ y(0) = \beta_1 \end{array} \right.$$

Vamos encontrar uma solução para a equação diferencial separando as variáveis:

$$\frac{dy(x)}{dx} = \frac{\beta_3}{\beta_2} y(x) (\beta_2 - y(x)) e^{-\beta_4 x} \Rightarrow \frac{dy(x)}{y(x) (\beta_2 - y(x))} = \frac{\beta_3}{\beta_2} e^{-\beta_4 x} dx \quad (45)$$

Note que existem $A, B \in \mathbb{R}$ tais que

$$\frac{1}{y(x)(\beta_2 - y(x))} = \frac{A}{y(x)} + \frac{B}{\beta_2 - y(x)} = \frac{(\beta_2 - y(x))A + By(x)}{y(x)(\beta_2 - y(x))} =$$

$$\frac{A\beta_2 + By(x) - Ay(x)}{y(x)(\beta_2 - y(x))} = \frac{A\beta_2 + (B - A)y(x)}{y(x)(\beta_2 - y(x))} \Rightarrow A\beta_2 = 1, B - A = 0$$

O que implica que

$$A = \frac{1}{\beta_2}, \quad (46)$$

$$B = A = \frac{1}{\beta_2} \quad (47)$$

Portanto, temos que

$$\frac{1}{\beta_2 y(x)} dy + \frac{1}{\beta_2 (\beta_2 - y(x))} dy = \frac{\beta_3}{\beta_2} e^{-\beta_4 x} dx,$$

ou seja,

$$\frac{1}{y(x)}dy + \frac{1}{\beta_2 - y(x)}dy = \beta_3 e^{-\beta_4 x} dx. \quad (48)$$

Integrando, obtemos

$$\ln|y(x)| - \ln|\beta_2 - y(x)| = -\frac{\beta_3}{\beta_4} e^{-\beta_4 x} + K, K \in \mathbb{R}$$

$$\Rightarrow \ln \left| \frac{y(x)}{\beta_2 - y(x)} \right| = -\frac{\beta_3}{\beta_4} e^{-\beta_4 x} + K$$

$$\Rightarrow \frac{y(x)}{\beta_2 - y(x)} = e^{\left(-\frac{\beta_3}{\beta_4} e^{-\beta_4 x} + K\right)} = e^{\left(-\frac{\beta_3}{\beta_4} e^{-\beta_4 x}\right)} e^K = e^{\left(-\frac{\beta_3}{\beta_4} e^{-\beta_4 x}\right)} C.$$

Podemos, então, dizer que

$$\begin{aligned} \frac{y(x)}{\beta_2 - y(x)} &= \frac{C}{e^{\left(\frac{\beta_3}{\beta_4} e^{-\beta_4 x}\right)}} \Rightarrow \\ y(x) e^{\left(\frac{\beta_3}{\beta_4} e^{-\beta_4 x}\right)} &= C(\beta_2 - y(x)) \Rightarrow \\ y(x) e^{\left(\frac{\beta_3}{\beta_4} e^{-\beta_4 x}\right)} &= C\beta_2 - Cy(x) \Rightarrow \\ y(x) e^{\left(\frac{\beta_3}{\beta_4} e^{-\beta_4 x}\right)} + Cy(x) &= C\beta_2 \Rightarrow \\ (e^{\left(\frac{\beta_3}{\beta_4} e^{-\beta_4 x}\right)} + C)y(x) &= C\beta_2 \Rightarrow \end{aligned}$$

$$y(x) = \frac{C\beta_2}{e^{\left(\frac{\beta_3}{\beta_4} e^{-\beta_4 x}\right)} + C} \quad (49)$$

Devemos ter de (I)

$$\lim_{x \rightarrow \infty} y(x) = \lim_{x \rightarrow \infty} \frac{C\beta_2}{e^{\left(\frac{\beta_3}{\beta_4} e^{-\beta_4 x}\right)} + C} = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1) e^{-\frac{\beta_3}{\beta_4}}}$$

Como $\beta_4 > 0$, temos que

$$\lim_{x \rightarrow \infty} \frac{C\beta_2}{e^{\left(\frac{\beta_3}{\beta_4}\right)e^{-\beta_4 x}} + C} = \frac{C\beta_2}{1 + C}.$$

Logo,

$$\begin{aligned} \frac{C\beta_2}{1 + C} &= \frac{\beta_1\beta_2}{\beta_1 + (\beta_2 - \beta_1)e^{-\frac{\beta_3}{\beta_4}}} \Leftrightarrow \\ C\beta_2[\beta_1 + (\beta_2 - \beta_1)e^{-\frac{\beta_3}{\beta_4}}] &= \beta_1\beta_2(1 + C) \Leftrightarrow \\ C\beta_2\beta_1 + C\beta_2^2e^{-\frac{\beta_3}{\beta_4}} - C\beta_2\beta_1e^{-\frac{\beta_3}{\beta_4}} &= \beta_2\beta_1 + C\beta_2\beta_1 \Leftrightarrow \\ C\beta_2e^{-\frac{\beta_3}{\beta_4}}(\beta_2 - \beta_1) &= \beta_2\beta_1 \Leftrightarrow \\ C &= \frac{\beta_1}{e^{-\frac{\beta_3}{\beta_4}}(\beta_2 - \beta_1)}. \end{aligned}$$

Voltando em (49), temos que

$$\begin{aligned} y(x) &= \frac{\frac{\beta_1}{e^{-\frac{\beta_3}{\beta_4}}(\beta_2 - \beta_1)}\beta_2}{e^{\left(-\frac{\beta_3}{\beta_4}\right)e^{-\beta_4 x}} + \frac{\beta_1}{e^{\left(-\frac{\beta_3}{\beta_4}\right)}(\beta_2 - \beta_1)}} \Leftrightarrow \\ y(x) &= \frac{\beta_1\beta_2}{\beta_1 + (\beta_2 - \beta_1)e^{-\frac{\beta_3}{\beta_4} + \frac{\beta_3}{\beta_4}e^{-\beta_4 x}}} \Leftrightarrow \\ y(x) &= \frac{\beta_1\beta_2}{\beta_1 + (\beta_2 - \beta_1)e^{-\frac{\beta_3}{\beta_4}(1 - e^{-\beta_4 x})}}. \end{aligned} \tag{50}$$

Para $x = 0$, temos

$$y(0) = \frac{\beta_1\beta_2}{\beta_1 + (\beta_2 - \beta_1)e^0} = \frac{\beta_1\beta_2}{\beta_1 + \beta_2 - \beta_1} = \frac{\beta_1\beta_2}{\beta_2} = \beta_1 \tag{51}$$

Portanto, (50) é solução do sistema (1).

Quando $\beta_4 < 0$ devemos buscar uma solução que satisfaça o seguinte sistema:

$$(2) \left\{ \begin{array}{l} \frac{dy(x)}{dx} = \beta_3 y(x) \left(1 - \frac{y(x)}{\beta_2} \right) e^{-\beta_4 x} \\ \lim_{x \rightarrow -\infty} y(x) = \frac{\beta_1 \beta_2}{\beta_1 + (\beta_2 - \beta_1) e^{-\frac{\beta_3}{\beta_4}}} \\ y(0) = \beta_1 \end{array} \right. \quad (52)$$

Analogamente ao caso em que $\beta_4 > 0$, temos que (50) é solução para o sistema (2).

A interpretação dos parâmetros do modelo Chanter pode ser feita similarmente a dos parâmetros dos modelos Logístico e Gompertz, pelo fato de suas equações diferenciais possuírem uma relação, descrita a seguir. A equação diferencial cuja solução nos fornece o modelo Logístico pode ser expressa por

$$\frac{dy(x)}{dx} = \beta_2 y(x) \left(1 - \frac{y(x)}{\beta_1} \right)$$

e a equação diferencial cuja solução nos fornece o modelo Gompertz pode ser expressa por

$$\frac{dy(x)}{dx} = \beta_2 e^{\beta_2 \beta_3} y(x) e^{-\beta_2 x}.$$

Observe que a equação diferencial do modelo Chanter dada por

$$y \frac{dy(x)}{dx} = \beta_3 y(x) e^{-\beta_4 x} \left(1 - \frac{y(x)}{\beta_2} \right)$$

é um misto entre as equações diferenciais dos modelos Logístico e Gompertz. De fato, o termo $\left(1 - \frac{y}{\beta_2} \right)$ depende do parâmetro β_2 assim como no modelo Logístico depende do parâmetro β_1 , o termo $e^{-\beta_4 x}$ depende da variável x e pode ser interpretado como diferenciação, desenvolvimento ou senescência como no modelo Gompertz.

4.2 Estudo de simulação

No cenário I, considerou-se uma amostra aleatória de tamanho 11 com 100 repetições gerada através do modelo Chanter para as variáveis comprimento e diâmetro do fruto do cacauero. Na Figura 11 é possível observar um histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas por meio do ajuste do modelo Chanter aos dados simulados de comprimento do fruto. Na mesma figura, a linha em vermelho representa o valor paramétrico utilizado na simulação dos dados e em vermelho tracejado a média das estimativas de cada um dos parâmetros.

Figura 11 - Cenário 1: Histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de comprimento do fruto em que linha em vermelho indica o valor paramétrico e a linha em vermelho tracejada indica a média das estimativas

Observe na Figura 11 que a média das estimativas e o verdadeiro valor do parâmetro estão muito próximos para todos os parâmetros avaliados. Os histogramas dos parâmetros β_1 e β_2 são mais simétricos, já os histogramas dos outros dois parâmetros β_3 e β_4 são assimétricos, à esquerda e à direita, respectivamente. Essa assimetria indica que os estimadores podem não ser normalmente distribuídas. Na Figura 12 está apresentado um box-plot dos vieses das estimativas de cada parâmetro obtidos a partir dos dados simulados de comprimento do fruto.

Observa-se nos gráficos box-plot da Figura 12 que apenas o parâmetro β_1 não apresenta vieses outliers, e aparentemente os vieses de todos os parâmetros possuem uma distribuição simétrica. β_2 é o parâmetro que possui vieses menores para uma amostra aleatória de tamanho 11 sendo que as maiores amplitudes dos vieses são observados para os parâmetros β_3 e β_4 .

Figura 12 - Cenário 1: Box-plot dos vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de comprimento do fruto

Na Figura 13 temos gráficos de dispersão dos AIC da simulação com os vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidos a partir dos dados simulados de comprimento do fruto.

Figura 13 - Cenário 1: Gráficos de dispersão dos AIC da simulação com os vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidos a partir dos dados simulados de comprimento do fruto

Observe na Figura 13 que não há relação proporcional nem inversamente proporcional entre o AIC e o viés das estimativas do parâmetro, para todos os parâmetros do modelo Chanter. É possível observar que para valores diferentes de AIC o vieses das estimativas do modelo Chanter permanecem constante.

Na Figura 14 é possível observar um histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas por meio do ajuste do modelo Chanter aos dados simulados para a variável diâmetro do fruto. Na mesma figura, a linha em vermelho representa o valor paramétrico utilizado na simulação dos dados e em vermelho tracejado a média das estimativas de cada um dos parâmetros.

Figura 14 - Cenário 1: Histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de diâmetro do fruto em que linha em vermelho indica o valor paramétrico e a linha em vermelho tracejada indica a média das estimativas

Observa-se que a média das estimativas e o verdadeiro valor do parâmetro estão muito próximos para os parâmetros β_1 , β_2 e β_4 . Os histogramas dos parâmetros β_1 , β_2 e β_4 são mais simétricos, já o histograma do parâmetro β_3 é assimétrico à esquerda. Essa assimetria indica que o estimador do parâmetro β_3 pode não ser normalmente distribuído.

Na Figura 15 está apresentado um box-plot dos vieses das estimativas de cada parâmetros obtidos a partir dos dados simulados de diâmetro do fruto.

Figura 15 - Cenário 1: Box-plot dos vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de diâmetro do fruto

Observa-se nos gráficos box-plot da Figura 15 que todos os parâmetros apresentam vieses outliers, e aparentemente os vieses de todos os parâmetros possuem uma distribuição simétrica. Os parâmetros β_1 , β_2 e β_3 apresentam as maiores amplitudes dos vieses. Como nos dados simulados para a variável comprimento, β_2 é o parâmetro que possui vieses menores para uma amostra aleatória de tamanho 11.

Na Figura 16 temos gráficos de dispersão entre o AIC dos 100 ajustes da simulação e os vieses das estimativas dos parâmetros do modelo para cada parâmetro do modelo Chanter para os dados simulados de diâmetro do fruto. Análogo a variável comprimento, observa-se que não há relação entre o AIC dos 100 ajustes da simulação e os vieses das estimativas dos parâmetros.

Figura 16 - Cenário 1: Gráficos de dispersão dos AIC da simulação com os vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidos a partir dos dados simulados de diâmetro do fruto

Considerando que a acurácia refere-se a pequenos erros totais, incluindo vício e variabilidade amostral, a acurácia dos estimadores foi avaliada por meio do erro quadrático médio, para isso foram calculados o valor médio da porcentagem dos vieses (V), o viés médio ($\bar{V}$) e o erro quadrático médio (EQM) dos estimadores dos parâmetros. Os resultados estão apresentados na Tabela 1:

Tabela 1 - Cenário I: Valor médio da porcentagem de viés, viés médio e erro quadrático médio dos estimadores dos parâmetros β_1 , β_2 , β_3 e β_4 do modelo Chanter para os dados simulados de comprimento e diâmetro de frutos de cacaueiro

	Comprimento			Diâmetro		
	V(%)	$\bar{V}$	EQM	V(%)	$\bar{V}$	EQM
β_1	2,41	$3,8 \cdot 10^{-2}(\text{ns})$	0,10	0,52	$2,9 \cdot 10^{-3}(\text{ns})$	$1,40 \cdot 10^{-2}$
β_2	0,03	$3,3 \cdot 10^{-3}(\text{ns})$	0,02	0,02	$1,9 \cdot 10^{-3}(\text{ns})$	$1,87 \cdot 10^{-2}$
β_3	-1,75	$-1,8 \cdot 10^{-4}(\text{ns})$	$1,66 \cdot 10^{-5}$	-20,45	$-2,6 \cdot 10^{-3}(p < 0,05)$	$1,63 \cdot 10^{-5}$
β_4	-5,02	$1,1 \cdot 10^{-3}(p < 0,05)$	$3,24 \cdot 10^{-5}$	-1,36	$2,1 \cdot 10^{-4}(\text{ns})$	$8,99 \cdot 10^{-6}$

Observe da Tabela 1 que o viés médio das estimativas dos parâmetros β_1 , β_2 e β_3 são nulos, isto é, não rejeita-se a hipótese $H_0 : \text{Viés} = 0$ ao nível de 5% de significância. Os valores de EQM para todos os parâmetros são pequenos, isto indica que os estimadores tem boa acurácia para os dados simulados de comprimento do fruto. Para a variável diâmetro do fruto, o viés médio das estimativas dos parâmetros β_1 , β_2 e β_4 são nulos e os valores de EQM para todos os parâmetros também são pequenos.

No segundo cenário, considerou-se uma amostra aleatória de tamanho 151 também com 100 repetições geradas através do modelo Chanter para comprimento e diâmetro do fruto do cacauieiro. Na Figura 17 é possível observar um histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas por meio do ajuste do modelo Chanter aos dados simulados de comprimento do fruto. Na mesma figura, a linha em vermelho representa o valor paramétrico utilizado na simulação dos dados e em vermelho tracejado a média das estimativas de cada um dos parâmetros.

Figura 17 - Cenário 2: Histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de comprimento do fruto em que linha em vermelho indica o valor paramétrico e a linha em vermelho tracejada indica a média das estimativas

Observe na Figura 17 que a média das estimativas e o verdadeiro valor do parâmetro estão muito próximos para todos os parâmetros avaliados. O histograma do parâmetro β_2 é o mais simétrico dentre os 4 parâmetros do modelo Chanter.

Na Figura 18 está apresentado um box-plot dos vieses das estimativas de cada parâmetro obtidos a partir dos dados simulados de comprimento do fruto. Observe-se nos gráficos box-plot da Figura 18 que todos os parâmetros apresentam vieses outliers, e aparentemente os vieses de todos os parâmetros possuem uma distribuição simétrica. Como no cenário 1, β_2 é o parâmetro que possui vieses menores mas agora para uma amostra aleatória de tamanho 151 sendo que a maior amplitude dos vieses é observado para o parâmetro β_3 .

Figura 18 - Cenário 2: Box-plot dos vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de comprimento do fruto

Na Figura 19 tem-se gráficos de dispersão para cada parâmetro do modelo Chanter para os dados simulados de comprimento do fruto com o intuito de investigar a relação entre o AIC dos 100 ajustes da simulação e os vieses das estimativas dos parâmetros do modelo em estudo.

Figura 19 - Cenário 2: Gráficos de dispersão dos AIC da simulação com os vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidos a partir dos dados simulados de comprimento do fruto

Observe na Figura 19 que não há relação proporcional nem inversamente proporcional entre o AIC e o viés das estimativas do parâmetro, para todos os parâmetros do modelo Chanter. É possível observar que para valores diferentes de AIC o vieses das estimativas do modelo Chanter permanecem constantes, mais evidente no cenário 2 cujos os vieses das estimativas dos parâmetros são muito pequenos.

Na Figura 20 é possível observar um histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas por meio do ajuste do modelo Chanter aos dados simulados para a variável diâmetro do fruto. Na mesma figura, a linha em vermelho representa o valor paramétrico utilizado na simulação dos dados e em vermelho tracejado a média das estimativas de cada um dos parâmetros.

Figura 20 - Cenário 2: Histograma das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de diâmetro do fruto em que linha em vermelho indica o valor paramétrico e a linha em vermelho tracejada indica a média das estimativas

Observa-se que a média das estimativas e o verdadeiro valor do parâmetro estão muito próximos para todos os parâmetros β_1 , β_2 , β_3 e β_4 . Os histogramas do parâmetro β_2 é o que possui maior simetria. A assimetria do histograma dos demais parâmetros do modelo indica que seus respectivos estimadores podem não ser normalmente distribuídos.

Na Figura 21 está apresentado um box-plot dos vieses das estimativas de cada parâmetros obtidos a partir dos dados simulados de diâmetro do fruto.

Figura 21 - Cenário 2: Box-plot dos vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidas a partir dos dados simulados de diâmetro do fruto

Observa-se nos gráficos box-plot da Figura 21 que apenas o parâmetro β_2 apresenta vieses outliers, e aparentemente os vieses de todos os parâmetros possuem uma distribuição simétrica. Os parâmetros β_1 , β_3 e β_4 apresentam amplitudes dos vieses semelhantes. Como nos dados simulados para a variável comprimento, β_2 é o parâmetro que possui vieses menores para uma amostra aleatória de tamanho 151.

Na Figura 22 temos gráficos de dispersão entre o AIC dos 100 ajustes da simulação e os vieses das estimativas dos parâmetros do modelo para cada parâmetro do modelo Chanter para os dados simulados de diâmetro do fruto. Análogo a variável comprimento, observa-se que não há relação entre o AIC dos 100 ajustes da simulação e os vieses das estimativas dos parâmetros.

Figura 22 - Cenário 2: Gráficos de dispersão dos AIC da simulação com os vieses das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 obtidos a partir dos dados simulados de diâmetro do fruto

Na Tabela 2 encontra-se valor médio da porcentagem dos vieses (V), o viés médio ($\bar{V}$) e o erro quadrático médio (EQM) dos estimadores dos parâmetros β_1 , β_2 , β_3 e β_4 do modelo Chanter para os dados simulados de comprimento e diâmetro de frutos de cacaueiro.

Tabela 2 - Cenário 2: Valor médio da porcentagem de viés, viés médio e erro quadrático médio dos estimadores dos parâmetros β_1 , β_2 , β_3 e β_4 do modelo Chanter para os dados simulados de comprimento e diâmetro de frutos de cacaueiro

	Comprimento			Diâmetro		
	V(%)	$\bar{V}$	EQM	V(%)	$\bar{V}$	EQM
β_1	-0,34	$-5,3 \cdot 10^{-3}(\text{ns})$	$8,09 \cdot 10^{-3}$	$7,38 \cdot 10^{-1}$	$4,2 \cdot 10^{-3}(\text{ns})$	$1,26 \cdot 10^{-3}$
β_2	0,06	$6,3 \cdot 10^{-3}(\text{ns})$	$1,43 \cdot 10^{-3}$	$-8,11 \cdot 10^{-4}$	$-6,87 \cdot 10^{-5}(\text{ns})$	$1,50 \cdot 10^{-3}$
β_3	-0,99	$-1 \cdot 10^{-4}(\text{ns})$	$1,18 \cdot 10^{-6}$	-23,32	$-3,03 \cdot 10^{-3} (p < 0,05)$	$9,86 \cdot 10^{-6}$
β_4	-10,07	$2,2 \cdot 10^{-3}(p < 0,05)$	$7,06 \cdot 10^{-6}$	-1,57	$2,4 \cdot 10^{-4} (p < 0,05)$	$7,80 \cdot 10^{-7}$

Observa-se da Tabela 2 que o viés médio das estimativas dos parâmetros β_1 ,

β_2 e β_3 são nulos para os dados simulados de comprimento, e para os dados simulados de diâmetro do fruto as estimativas dos parâmetros β_1 , β_2 têm viés médio nulo. Os valores de EQM para todos os parâmetros são pequenos, isto indica que os estimadores tem boa acurácia para os dados simulados de comprimento e diâmetro do fruto.

Ambos os cenários apresentaram menor valor da porcentagem de viés, do viés médio e do erro quadrático médio dos estimadores para o parâmetro β_2 , indicando com base no estudo da função Chanter que o estimador do parâmetro que representa o tamanho assintótico do fruto é o que tem melhor acurácia para os dados simulados de comprimento e diâmetro do fruto.

4.3 Estudo dos dados reais

É apresentado, na Figura 23 o diagrama de dispersão dos dados de comprimento e diâmetro do fruto do cacau. Serão apresentadas as análises realizadas para o ajuste dos modelos Chanter, Logístico e Gompertz a esses dados.

Figura 23 - Gráficos de dispersão dos dados de comprimento e diâmetro de cacau ao longo do tempo

É possível observar por meio da Figura 23 que não houveram valores discrepantes em cada coleta para as medidas de comprimento e diâmetro do fruto do cacau. A média dos dados ao longo dos 180 dias do experimento e ajuste das curvas dos modelos Logístico, Gompertz e Chanter para a média dos dados de comprimento e diâmetro estão representados na Figura 24. Temos indícios que o comportamento dos dados aproxima-se de uma curva sigmoidal. Observa-se também que o comprimento do fruto de cacau estabiliza-se

em aproximadamente 11 cm, já o diâmetro em aproximadamente 10 cm e que o crescimento mais acentuado ocorre até aproximadamente 90 dias após a primeira coleta, depois desse período há uma desaceleração do crescimento.

Figura 24 - Médias e ajuste dos modelos Logístico, Gompertz e Chanter aos dados de: (a) comprimento e (b) diâmetro de cacau ao longo do tempo

Nas Tabelas 3 e 4 são apresentados os valores das estimativas dos parâmetros β_1 , β_2 , β_3 e β_4 .

Tabela 3 - Estimativas dos parâmetros (erros-padrão) e p-valor do teste de Shapiro-Wilk dos modelos Chanter, Logístico e Gompertz aos dados de comprimento de frutos de cacauero

	Comprimento		
	Chanter	Logístico	Gompertz
β_1	1,56 (0,315)	10,89 (0,282)	11,26 (0,509)
β_2	10,55 (0,124)	-3,15 (0,324)	0,03 (0,004)
β_3	0,01 (0,004)	0,04 (0,005)	54,93 (3,208)
β_4	-0,02 (0,005)		
p	0,4945	0,8209	0,9978

Tabela 4 - Estimativas dos parâmetros (erros-padrão) e p-valor do teste de Shapiro-Wilk dos modelos Chanter, Logístico e Gompertz aos dados de diâmetro de frutos de cacauero

	Diâmetro		
	Chanter	Logístico	Gompertz
β_1	0,57 (0,131)	8,84 (0,265)	9,47 (0,669)
β_2	8,47 (0,102)	-4,31 (0,380)	0,03 (0,005)
β_3	0,01 (0,004)	0,04 (0,004)	79,11 (4,257)
β_4	-0,01 (0,003)		
p	0,4564	0,1372	0,1817

Para testar a normalidade dos resíduos, foi utilizado o teste de Shapiro-Wilk, aplicado nos resíduos dos modelos não lineares Chanter, Logístico e Gompertz. Como o p-valor do teste foi maior que o nível de 5% de significância, não rejeitamos a hipótese de normalidade dos resíduos e, ou seja, conclui-se que os erros são normalmente distribuídos.

Nas Figuras 25 e 26 estão apresentados os gráficos dos resíduos padronizados para as medidas de comprimento e diâmetro do fruto do cacauero.

Figura 25 - Gráficos dos resíduos padronizados para a medida de comprimento do cacau para os modelos Logístico, Gompertz e Chanter respectivamente

Figura 26 - Gráficos dos resíduos padronizados para a medida de diâmetro do cacau para os modelos Logístico, Gompertz e Chanter respectivamente

Através das Figuras 25 e 26, observa-se o comportamento dos resíduos padronizados ao longo das medições. A princípio, não se percebe nenhuma espécie de padrão dos resíduos nos gráficos.

Nas Figuras 27 e 28 estão apresentados os gráficos Normal de probabilidades com envelopes de confiança simulados para as medidas de comprimento e diâmetro do fruto do cacaueiro.

Figura 27 - Gráficos Normal de probabilidade para a medida de comprimento do cacau para os modelos Logístico, Gompertz e Chanter respectivamente

Figura 28 - Gráficos Normal de probabilidade para a medida de diâmetro do cacau para os modelos Logístico, Gompertz e Chanter respectivamente

Observe que os gráficos Normal de probabilidades apresentados nas Figuras 27 e 28, indicam que os resíduos dos modelos Logístico, Gompertz e Chanter respectivamente, apresentam distribuição normal.

A fim de comparar os ajustes dos três modelos e diagnosticar qual é o mais adequado para a descrição do fenômeno, já que todos se ajustam bem aos dados, foram obtidos os valores pelos critérios que estão apresentados na Tabelas 5.

Tabela 5 - Avaliadores da qualidade de ajuste dos modelos Chanter, Logístico e Gompertz aos dados de comprimento e diâmetro de frutos de cacaueiro

	Comp.			Diam.		
	AIC	BIC	DPR	AIC	BIC	DPR
Chanter	6,71	8,70	0,26	-2,6	-0,6	0,17
Logístico	19,06	20,65	0,47	12,45	14,04	0,35
Gompertz	25,72	27,32	0,63	22,3	23,89	0,54

Com base nos critérios apresentados na Tabela 5, é possível indicar que o modelo mais adequado para o ajuste dos dados é o modelo Chanter, pois os valores de todos os critérios avaliados para esse modelo foram inferiores quando comparados aos modelos Logístico e Gompertz.

Para o cálculo do critério peso de Akaike, é necessário obter a diferença entre os valores do AICc dos modelos Chanter, Logístico e Gompertz, como na Tabela 6.

Tabela 6 - Diferença entre os valores de AICc (valor do Peso de Akaike) dos modelos Chanter (C), Logístico (L) e Gompertz (G) aos dados de comprimento e diâmetro de frutos de cacaueteiro

	Comprimento	Diâmetro
(L - C)	7,01 (0,0292)	9,72 (0,0077)
(G - C)	13,68 (0,0011)	19,57 (0,0001)

Tem-se na Tabela 6 que para os dados de comprimento o modelo Logístico tem uma probabilidade de 2,92% de ser correto comparado ao modelo Chanter e o modelo Gompertz tem uma probabilidade de 0,11% de ser correto comparado ao modelo Chanter. Para os dados de diâmetro o modelo Logístico tem uma probabilidade de 0,77% de ser correto comparado ao modelo Chanter e o modelo Gompertz tem uma probabilidade de 0,01% de ser correto comparado ao modelo Chanter. Com base na Tabela 6, o modelo Chanter mostrou-se o modelo com maior chance de ser o correto comparado com os modelos Logístico e Gompertz para os dados de comprimento e diâmetro do fruto do cacaueteiro. Podemos concluir que tanto para os dados de comprimento quanto para os dados de diâmetro, o modelo Chanter apresentou maior porcentagem de chance de ser o modelo correto comparado com os modelos Logístico e Gompertz.

Os valores calculados para a medida de não linearidade Curvatura de Bates e Watts encontram-se na Tabela 7. Nesta tabela apresentamos a medida de não linearidade intrínseca (IN) e a medida de não linearidade devido aos efeitos dos parâmetros (EP), neste trabalho o interesse é na medida de não linearidade intrínseca, que mede a curvatura do espaço de estimação no espaço amostral. Para um modelo não linear, o espaço de estimação é curvelíneo, e (IN) mede a extensão dessa curvatura.

Observa-se que tanto para os dados de comprimento quanto para os dados de diâmetro a medida de não linearidade intrínseca é menor para o modelo Logístico, isto é, apresenta o comportamento mais próximo do comportamento linear. Para o comprimento o modelo Chanter apresenta um comportamento mais distante do comportamento linear, essa não linearidade é devida, principalmente, ao efeito de parâmetros. Já para o diâmetro, o modelo Gompertz apresenta um comportamento mais distante do comportamento linear.

Tabela 7 - Valores da medida de não linearidade curvatura de Bates e Watts para os modelos Chanter, Logístico e Gompertz aplicados aos dados de comprimento e diâmetro de frutos de cacauero

	Comprimento		Diâmetro	
	IN	EP	IN	EP
Chanter	0,2469	7,2107	0,1695	7,1212
Logístico	0,1562	0,5193	0,151	0,5656
Gompertz	0,1771	0,9612	0,2923	1,4022

A medida de não linearidade vício de Box nos mostra a importância em se avaliar os vícios pois indicam qual ou quais parâmetros do modelo são os maiores responsáveis pelo comportamento distante do comportamento linear. Encontra-se na Tabela 8 os valores dos vícios para os modelos Logístico e Gompertz.

Tabela 8 - Valores em porcentagem da medida de não linearidade vício de Box para os modelos Chanter, Logístico e Gompertz aos dados de comprimento e diâmetro de frutos de cacauero

	Comprimento			Diâmetro		
	Chanter	Logístico	Gompertz	Chanter	Logístico	Gompertz
β_1	0,3021	0,1172	0,3708	0,6574	0,1358	0,6878
β_2	0,0341	0,6099	1,0009	0,0411	0,5452	1,4874
β_3	3,2244	0,6041	0,3058	1,9710	0,5457	0,5167
β_4	1,0590			0,4269		

Podemos indicar da Tabela 8 que o parâmetro β_2 do modelo Gompertz e os parâmetros β_3 e β_4 do modelo Chanter são os responsáveis pelo comportamento distante do comportamento linear, tanto para o comprimento quanto para o diâmetro, pois apresenta vício maior que 1%. Esses dois modelos, Chanter e Gompertz, possuem em suas equações uma exponencial negativa que contém os parâmetros β_3 , β_4 e β_2 respectivamente, em função do tempo e o resultado obtido através da medida de não linearidade Vício de Box reforça a ideia que são esses parâmetros os responsáveis pelo comportamento flexível destes modelos.

Com base nos avaliadores de qualidade de ajuste apresentados acima, temos indícios de que o modelo Chanter é o melhor para o ajuste dos dados reais estudados neste

trabalho, sendo assim, a equação de predição do modelo Chanter para o comprimento do fruto do cacaueiro fica:

$$\hat{y}_i = \frac{16,46}{1,56 + 8,99e^{0,5(1-e^{0,02x_i})}}. \quad (53)$$

E a equação de predição do modelo Chanter para o diâmetro do fruto do cacaueiro fica:

$$\hat{y}_i = \frac{4,83}{0,57 + 7,9e^{(1-e^{0,01x_i})}}. \quad (54)$$

Os gráficos destas curvas estão apresentados na Figura 29. Observe que as curvas estimadas apresentam mudança de concavidade, ou seja, o crescimento do fruto passa de uma velocidade positiva para uma velocidade negativa. Com as estimativas dos parâmetros é possível identificar o ponto de inflexão destas curvas que está apresentado na Figura 30.

Figura 29 - Gráficos de dispersão, médias e ajuste do modelo Chanter aos dados de (a) comprimento e (b) diâmetro do fruto do cacaueiro ao longo do tempo

Figura 30 - Médias, ajuste do modelo Chanter aos dados de (a) comprimento e (b) diâmetro do fruto do cacaueiro ao longo do tempo e ponto de inflexão

Para o comprimento do fruto do cacaueiro a mudança de concavidade ocorre no ponto $(83, 39; 7, 08)$, isso nos diz que o crescimento do fruto tem uma velocidade positiva até o 83º dia e depois passa para uma velocidade negativa. Para o diâmetro do fruto do cacaueiro a mudança de concavidade ocorre no ponto $(100, 80; 5, 29)$, isso nos diz que o crescimento do fruto tem uma velocidade positiva até o 100º dia e depois passa para uma velocidade negativa.

5 CONCLUSÃO

Os parâmetros β_1 e β_2 possuem interpretação prática, sendo β_1 o intercepto, indica o tamanho do fruto no dia zero, e β_2 é a assíntota superior á direita, indica tamanho máximo assintótico do fruto. Diante disso, realizar estudos utilizando o modelo Chanter que explorem a interpretação prática dos outros dois parâmetros seria pertinente em trabalhos futuros.

Os modelos Logístico, Gompertz e Chanter se mostraram adequados para descrever o comportamento do comprimento e diâmetro do fruto do cacaueiro ao longo do tempo, entretanto o modelo Chanter foi o melhor.

O modelo Chanter possui uma maior flexibilidade no ajuste aos dados, isso ocorre devido a vários fatores, sendo um deles a presença de um parâmetro a mais quando comparado aos modelos Logístico e Gompertz.

Foi obtida uma estimativa de aproximadamente 10,55 cm para comprimento e 8,47 cm para diâmetro do parâmetro β_2 representando a medida assintótica do fruto. O parâmetro β_1 representa o tamanho do fruto no dia 0, ou seja, 30 dias antes da polinização, o comprimento do fruto era 1,56 cm e o diâmetro era 0,57 cm. Já os outros dois parâmetros não apresentam interpretação prática imediata.

REFERÊNCIAS

- Akaike, H. A. *A new look at the statistical model identification*. IEEE Transation on Automatic Control. New York, v. 19, p. 716 - 723, 1974.
- Alexandre, R. S.; Chagas, K.; Marques, H. I. P.; Costa, P. R.; Filho, J. C. *Caracterização de frutos de clones de cacaueiros na região litorânea de São Mateus, ES*. Revista Brasileira de Engenharia Agrícola e Ambiental, 2015.
- Alves, M. C. G. P.; Silva, N. N. *Métodos de estimação de variância em amostras provenientes de inquéritos domiciliares*. Revista Saúde Pública, 2007.
- Bates, D.M.; Watts, D. G. *Relative Curvature Measures of Nonlinearity* Journal of the Royal statistical Society. Series B(Methodological). London, v. 42, p. 1 - 25, 1980.
- Box, M. L. *Bias in nonlinear estimation*. Journal of the Royal statistical Society. Series B(Methodological). london, v. 33, n. 2, p. 171 - 201, 1971.
- Brito, I. C.; Silva, C. P. *Medidas biométricas do fruto do cacaueiro durante seu desenvolvimento*, 1983.
- Cardiac Safety Consultants, <<http://cardiacsafetyconsultants.com/consultants.php>> Acessado em 29 de janeiro de 2017.
- Chanter, D. O.; Thornley, J. H. M. *Mycelial Growth and the Initiation and Growth of Sporophores in the Mushroom Crop: a Mathematical Model*, 1977.
- Cox, D. R.; Snell, E. J. *A General Definition of Residuals*. Journal of the Royal statistical Society. Series B(Methodological). london, v. 20, n.2, p. 248 - 275, 1968.
- Dempster, A. P.; Laird, N.M. e Rubin, D.B. *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society, B, 39, p. 1-22, 1977.
- Domingues, J. S.; *Análise do Modelo de Gompertz no crescimento de tumores sólidos e inserção de um fator de tratamento* , 2011.
- Draper, N. R.; Smith, H. *Applied Regression Analysis*. ISBN: 978-0-471-17082-2, 3 ed., 1998.
- France, J.; Thornley, J. H. M. *Mathematical models in agriculture* London: Butterworks, 1984. p.335.
- Freitas, A. R. *Curvas de crescimento na produção animal*, 2005.
- Gray, V. M. *A Comparison of Two Approaches for Modelling Cassava (Manihot esculenta Crantz.) Crop Growth*, 2000.
- Guidorizzi, H. L. *Um curso de cálculo*, Vol. 1 - Ed. 1997.
- Laird, A. K. *et al*, *Dynamics of relative growth.*, 1965.
- Mazucheli, J.; Achcar, J. A. *Algumas considerações em regressão não linear*, 2002.

- Migault, V.; Combes, D.; Barre, P.; Gueye, B.; Louarn, G.; Escobar-Gutiérrez, A.J. *Improved modelling of ryegrass foliar growth*. In Proceedings of the Plant Growth Modeling, Simulation, Visualization and Applications PMA12, Shanghai, China, 31 October–3 November 2012; IEEE: Beijing, China, 2012; pp. 282–288.
- Mischan, M. M.; Pinho, S. Z. *Modelos não lineares*. 1. ed. - São Paulo: Cultura Acadêmica, 2014.
- Mazucheli, J.; Achcar, J. A. *Algumas considerações em regressão não linear*, 2002.
- Mendes, P. N.; *Ajuste dos Modelos Gompertz e Logístico na descrição das curvas de crescimento de tomates: uma abordagem bayesiana*, Tese UFPA, 2011.
- Motulsky, H.; Christopoulos, A. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. 4. ed. San Diego, CA: GraphPad Software, 2003.
- Muniz, J. A.; Nascimento, M. S.; Fernandes, T. J. *Modelos não lineares na descrição do crescimento de frutos de cacau com violações de pressupostos*, 2017.
- Nelder, J.; *The fitting of a generalization of the logistic curve*. Biometrics, 1961.
- Pinheiro, J. C.; Bates, D. M. *Mixed-Effects Models in S-PLUS*. New York: Springer-Verlag, 2000. p.528.
- Perre, M. A.; Lima, E. A.; Asanome, C. R.; Ludovico, J.; *Um estudo de simulação para avaliação do comportamento do modelo de Mitscherlich*. Semina, 1989.
- Ratkowsky, D. A. *Nonlinear regression modeling*. Dekker, New York., 1983.
- Schabenberger, O.; Pierce, F.J. *Contemporary statistical models for the plant and soil sciences*. Boca Raton: CRC, 2002. p.753.
- Seber, G.; Wild, C. *Nonlinear Regression*. [S.l.]: Wiley, 2003. (Wiley Series in Probability and Statistics). ISBN 9780471471356.
- Segarra, J.; Jeger, M.J.; Bosch, F. van den. *Epidemic dynamics and patterns of plant diseases*. The American Phytopathological Society, v. 91, n. 10, out. 2001.
- Schwarz, G. *Estimating the dimension of a model*. Annals of Statistics. Hayward, v.6, n.2, p. 461–464, 1978.
- Smyth, G. K. *Nonlinear regression*. In: Encyclopedia of environmetrics. Chichester: J. Wiley, 2002, v. 3, p. 1405 - 1411.
- Souza, G. S. *Introdução aos modelos de regressão linear e não linear* Brasília: Embrapa/SEA, 1998.
- Tavoni, R.; Oliveira, R. Z. G. *Os modelos de crescimento populacional de Malthus e Verhulst - uma motivação para o ensino de logaritmos e exponenciais* 2013.
- Werker, A. R.; Jaggard, K. W. *Modelling Asymmetrical Growth Curves that Rise and then Fall : Applications to Foliage Dynamics of Sugar Beet (Beta vulgaris L.)* IACR Broom's Barn, Higham, Bury St Edmunds, Suffolk IP28 6NP, UK, 1997.

Zill, D. G. *Equações Diferenciais com Aplicações em Modelagem*. vol. 2, 2001.

Zeviani², W. M. et al. *Modelos não-lineares para a liberação de potássio de esterco animais em latossolos*. Ciência Rural, Santa Maria, v. 42, n. 10, p. 1789-1796, out. 2012.

APÊNDICES

Modelo Logístico

Considere a seguinte equação diferencial

$$\frac{dy(x)}{dx} = \beta_2 y(x) \left(1 - \frac{y(x)}{\beta_1} \right)$$

Vamos encontrar uma solução para a equação diferencial separando as variáveis:

$$\frac{dy(x)}{dx} = \frac{\beta_2}{\beta_1} y(x)(\beta_1 - y(x)) \Rightarrow \frac{dy(x)}{y(x)(\beta_1 - y(x))} = \frac{\beta_2}{\beta_1} dx$$

Note que existem $A, B \in \mathbb{R}$ tais que

$$\frac{1}{y(x)(\beta_1 - y(x))} = \frac{A}{y(x)} + \frac{B}{\beta_1 - y(x)} = \frac{(\beta_1 - y(x))A + By(x)}{y(x)(\beta_1 - y(x))} =$$

$$\frac{A\beta_1 + By(x) - Ay(x)}{y(x)(\beta_1 - y(x))} = \frac{A\beta_1 + (B - A)y(x)}{y(x)(\beta_1 - y(x))} \Rightarrow A\beta_1 = 1, B - A = 0$$

O que implica que

$$A = \frac{1}{\beta_1}, \tag{55}$$

$$B = A = \frac{1}{\beta_1} \tag{56}$$

Portanto, temos que

$$\frac{1}{\beta_1 y(x)} dy + \frac{1}{\beta_1(\beta_1 - y(x))} dy = \frac{\beta_2}{\beta_1} dx,$$

ou seja,

$$\frac{1}{y(x)} dy + \frac{1}{\beta_1 - y(x)} dy = \beta_2 dx. \tag{57}$$

Integrando, obtemos

$$\ln|y(x)| - \ln|\beta_1 - y(x)| = \beta_2 x + \beta_3, \beta_3 \in \mathbb{R}$$

$$\Rightarrow \ln \left| \frac{y(x)}{\beta_1 - y(x)} \right| = \beta_2 x + \beta_3$$

$$\Rightarrow \frac{y(x)}{\beta_1 - y(x)} = e^{\beta_2 x + \beta_3}$$

Podemos, então, dizer que

$$\begin{aligned}
 y(x) &= e^{\beta_2 x + \beta_3} (\beta_1 - y(x)) \Rightarrow \\
 y(x) &= e^{\beta_2 x + \beta_3} \beta_1 - e^{\beta_2 x + \beta_3} y(x) \Rightarrow \\
 y(x) + e^{\beta_2 x + \beta_3} y(x) &= e^{\beta_2 x + \beta_3} \beta_1 \Rightarrow \\
 (1 + e^{\beta_2 x + \beta_3}) y(x) &= e^{\beta_2 x + \beta_3} \beta_1 \Rightarrow \\
 y(x) &= \frac{e^{\beta_2 x + \beta_3} \beta_1}{1 + e^{\beta_2 x + \beta_3}}.
 \end{aligned}$$

Dividindo por $e^{\beta_2 x + \beta_3}$, obtemos

$$y(x) = \frac{\beta_1}{1 + e^{-\beta_3 - \beta_2 x}}.$$

□

Modelo Gompertz

Considere a seguinte equação diferencial

$$\frac{y'(x)}{y(x)} = \beta_2 e^{(-\beta_2(x-\beta_3))}$$

Multiplicando ambos os lados da igualdade por $y(x)$ obtemos

$$y'(x) = \frac{\beta_2}{e^{\beta_2(-\beta_3+x)}} y(x)$$

Substituindo $y'(x)$ por $\frac{dy}{dx}$ e $y(x)$ por y temos

$$\frac{dy}{dx} = \frac{\beta_2}{e^{\beta_2(-\beta_3+x)}} y. \quad (58)$$

Uma equação diferencial da forma $\frac{dy}{dx} = \frac{g(x)}{h(y)}$ é chamada separável (ZILL, 2001). Observe que (58) é uma equação separável. Então multiplicando a igualdade em cruz e separando de um lado da igualdade o x e do outro lado o y tem-se:

$$e^{\beta_2(-\beta_3+x)} dy = \beta_2 y dx \Leftrightarrow e^{-\beta_2 \beta_3} e^{\beta_2 x} dy = \beta_2 y dx$$

$$\Updownarrow$$

$$\frac{e^{-\beta_2 \beta_3}}{y} dy = \beta_2 e^{-\beta_2 x} dx$$

Integrando em ambos os lados obtém-se

$$\int \frac{e^{-\beta_2 \beta_3}}{y} dy = \int \beta_2 e^{-\beta_2 x} dx \Leftrightarrow e^{-\beta_2 \beta_3} \ln |y| + c_1 = \beta_2 \frac{e^{-\beta_2 x}}{-\beta_2} + c_2$$

em que c_1 e c_2 são constantes.

Tome $c = c_2 - c_1$. Assim, $e^{-\beta_2\beta_3} \ln |y| = \beta_2 \frac{e^{-\beta_2 x}}{-\beta_2} + c$. Isolando y tem-se

$$\ln |y| = \frac{-e^{-\beta_2 x} + c}{e^{-\beta_2\beta_3}} \Leftrightarrow \ln |y| = -e^{-\beta_2 x + \beta_2\beta_3} + \frac{c}{e^{-\beta_2\beta_3}} \Leftrightarrow \ln |y| = -e^{-\beta_2 x + \beta_2\beta_3} + ce^{\beta_2\beta_3}$$

Aplicando exponencial em ambos os lados tem-se $y = e^{-e^{-\beta_2 x + \beta_2\beta_3} + ce^{\beta_2\beta_3}}$. Isto é, $y = e^{(-e^{-\beta_2 x + \beta_2\beta_3})e^{ce^{\beta_2\beta_3}}}$.

Chamando $e^{ce^{\beta_2\beta_3}}$ de β_1 obtém-se a equação do modelo Gompertz.

$$y(x) = \beta_1 e^{(-e^{(-\beta_2(x-\beta_3))})}$$

□

ANEXOS

```
#####
##### ROTINA NO R #####
#####

#ESTUDO DE SIMULACAO

# Cenario 1: n=11

#COMPRIMENTO

#Valores parametricos
>a=1.56
>b=10.55
>c=0.01
>d=-0.02
>s=0.26
>x=seq(2,12,1)*15
>n = length(x)

#Simulacao
>int=100
>theta=matrix(numeric(0),int,5)

vetor_aic <- c()

for(i in 1:int){
  set.seed(2000*i);
  y=((a*b)/(a+(b-a)*exp(-c/d*(1 - exp(-d*x))))) + rnorm(n,0,s)

  a11=2; b11=10; c11=0.01; d11=-0.03 #Valores iniciais para os param.

  try(ajuste1<-nls(y~((A1*B1)/(A1+(B1-A1)*exp((-C1/D1)*(1-exp(-D1*x))))),
    start=c(A1=a11, B1=b11, C1=c11, D1=d11 ),TRUE)
  fit1=summary(ajuste1)
  param1=c(fit1$coef[,1],fit1$sigma)
  theta[i,]=param1

  vetor_aic[i] <- AIC(ajuste1)
}

#Matriz com as estimativas dos 4 parametros e do sigma
>theta.new1=theta[!duplicated(theta),];theta.new1
>medial = colMeans(theta.new1,na.rm = T);medial
>var1 = diag(var(theta.new1,na.rm = T));var1
>vlrverd = c(1.56461, 10.54930, 0.01018, -0.02219, 0.2612);vlrverd
```

```

#Vieses de cada parametro
>vies1= (theta.new1[,1] - vlrverd[1])/ vlrverd[1]; vies1
>vies2= (theta.new1[,2] - vlrverd[2])/ vlrverd[2]; vies2
>vies3= (theta.new1[,3] - vlrverd[3])/ vlrverd[3]; vies3
>vies4= (theta.new1[,4] - vlrverd[4])/ vlrverd[4]; vies4
>mediavies = c(mean(vies1), mean(vies2), mean(vies3), mean(vies4));
mediavies
>100*mediavies
#ou
>vies = ( media1 - vlrverd)/ vlrverd; vies
>100*vies
#teste
>zcal1= ( media1[1] - vlrverd[1])/ sqrt(var1[1]/int); zcal1
>zcal2= ( media1[2] - vlrverd[2])/ sqrt(var1[2]/int); zcal2
>zcal3= ( media1[3] - vlrverd[3])/ sqrt(var1[3]/int); zcal3
>zcal4= ( media1[4] - vlrverd[4])/ sqrt(var1[4]/int); zcal4
>ztab = 1.96
>abs(zcal1) > ztab
>abs(zcal2) > ztab
>abs(zcal3) > ztab
>abs(zcal4) > ztab

>beta1 = vies1
>beta2 = vies2
>beta3 = vies3
>beta4 = vies4

#Boxplot dos vieses
>vies_d = data.frame(beta1, beta2, beta3, beta4); vies_d
>boxplot(vies_d, xlab = "Parametros", ylab = "Vies", axes=T)

#Histograma
>par(mfrow=c(2,2))
>hist(theta.new1[,1], main = "", xlab = "Amostras beta1",
ylab="Frequência")
>abline(v=1.5646, col = "red", lwd = 2)
>abline(v= 1.6023, col ="red", lwd = 2, lty=2)

>hist(theta.new1[,2], main = " ", xlab = "Amostras beta2",
ylab="Frequência")
>abline(v=10.5493, col = "red", lwd = 2)
>abline(v= 10.5526, col ="red", lwd = 2, lty=2)

>hist(theta.new1[,3], main = "", xlab = "Amostras beta3",
ylab="Frequência")
>abline(v=0.0101, col = "red", lwd = 2)

```

```

>abline(v= 0.0100, col ="red",lwd = 2, lty=2)

>hist(theta.new1[,4],main = "", xlab = "Amostras beta4",
ylab="Frequência")
>abline(v=-0.0221, col = "red", lwd = 2)
>abline(v= -0.0210, col ="red",lwd = 2, lty=2)

#EQM
>EQM1 = var1[1] + (media1[1] - vlrverd[1])^2; EQM1
>EQM2 = var1[2] + (media1[2] - vlrverd[2])^2; EQM2
>EQM3 = var1[3] + (media1[3] - vlrverd[3])^2; EQM3
>EQM4 = var1[4] + (media1[4] - vlrverd[4])^2; EQM4
>EQM= c(EQM1, EQM2, EQM3, EQM4); EQM

#AIC e vies
>par(mfrow=c(2,2))
>plot(vetor_aic, vies1,main = "", xlab = "AIC ",
ylab="Vies beta1", ylim=c(-1, 1.5))
>plot(vetor_aic, vies2,main = "", xlab = "AIC ",
ylab="Vies beta2",ylim=c(-1, 1.5))
>plot(vetor_aic, vies3,main = "", xlab = "AIC ",
ylab="Vies beta3",ylim=c(-1, 1.5))
>plot(vetor_aic, vies4,main = "", xlab = "AIC ",
ylab="Vies beta4",ylim=c(-1, 1.5))

#DIAMETRO

#Valores parametricos
>a1=0.57
>b1=8.47
>c1=0.01
>d1=-0.015
>s1=0.17
>x=seq(2,12,1)*15
>n = length(x)

#Simulacao
>int=100
>theta1=matrix(numeric(0),int,5)

>vetor_aic <- c()

>for(i in 1:int){
  set.seed(2000*i);
  y=((a1*b1)/(a1+(b1-a1)*exp(-c1/d1*(1 - exp(-d1*x))))) + rnorm(n,0,s1)

  a11=1; b11=8; c11=0.01; d11=-0.02 #Valores iniciais para os parametros

```

```

try(
ajuste11<-nls(y~((A1*B1)/(A1+(B1-A1)*exp((-C1/D1)*(1-exp(-D1*x))))),
start=c(A1=a11, B1=b11, C1=c11, D1=d11 ),TRUE)
fit11=summary(ajuste11)
param11=c(fit11$coef[,1],fit11$sigma)
theta1[i,]=param11
vetor_aic[i] <- AIC(ajuste11)
}

#Matriz com as estimativas dos 4 parametros e do sigma
>theta.new11=theta1[!duplicated(theta1), ];theta.new11
>media11 = colMeans(theta.new11,na.rm = T);media11
>var11 = diag(var(theta.new11,na.rm = T));var11
>vlrverd1 = c(0.568917, 8.4738, 0.0130, -0.0153, 0.1777);vlrverd1

#Vieses de cada parametro
>vies11= (theta.new11[,1] - vlrverd1[1])/vlrverd1[1];vies11
>vies21= (theta.new11[,2] - vlrverd1[2])/vlrverd1[2];vies21
>vies31= (theta.new11[,3] - vlrverd1[3])/vlrverd1[3];vies31
>vies41= (theta.new11[,4] - vlrverd1[4])/vlrverd1[4];vies41
>mediavies1 = c(mean(vies11), mean(vies21), mean(vies31), mean(vies41));
>mediavies1
>100*mediavies1
#ou
>vies1 = ( media11 - vlrverd1)/vlrverd1; vies1
>100*vies1

#teste
>zcal1d= ( media11[1] - vlrverd1[1])/sqrt(var11[1]/int);zcal1d
>zcal2d= ( media11[2] - vlrverd1[2])/sqrt(var11[2]/int);zcal2d
>zcal3d= ( media11[3] - vlrverd1[3])/sqrt(var11[3]/int);zcal3d
>zcal4d= ( media11[4] - vlrverd1[4])/sqrt(var11[4]/int);zcal4d
>ztab = 1.96
>abs(zcal1d) > ztab
>abs(zcal2d) > ztab
>abs(zcal3d) > ztab
>abs(zcal4d) > ztab

>beta1 = vies11
>beta2 = vies21
>beta3 = vies31
>beta4 = vies41

#boxplot dos vieses
>vies_d1 = data.frame(beta1, beta2, beta3, beta4);vies_d1
>boxplot(vies_d1,xlab ="Parametros", ylab="Vies", axes=T)

```

```

#Histograma boxplot dos vieses
>par(mfrow=c(2,2))
>hist(theta.new11[,1],main = "", xlab = "Amostras beta1",
ylab="Frequencia")
>abline(v=0.5689, col = "red", lwd = 2)
>abline(v= 0.5704, col ="red",lwd = 2, lty=2)

>hist(theta.new11[,2],main = " ", xlab = "Amostras beta2",
ylab="Frequencia")
>abline(v=8.4738, col = "red", lwd = 2)
>abline(v= 8.4911, col ="red",lwd = 2, lty=2)

>hist(theta.new11[,3],main = "", xlab = "Amostras beta3",
ylab="Frequencia")
>abline(v=0.0130, col = "red", lwd = 2)
>abline(v= 0.0105, col ="red",lwd = 2, lty=2)

>hist(theta.new11[,4],main = "", xlab = "Amostras beta4",
ylab="Frequencia")
>abline(v=-0.0153, col = "red", lwd = 2)
>abline(v= -0.0150, col ="red",lwd = 2, lty=2)

#EQM
>EQM11 = var11[1] + (media11[1] - vlrverd1[1])^2; EQM11
>EQM21 = var11[2] + (media11[2] - vlrverd1[2])^2; EQM21
>EQM31 = var11[3] + (media11[3] - vlrverd1[3])^2; EQM31
>EQM41 = var11[4] + (media11[4] - vlrverd1[4])^2; EQM41
>EQM1= c(EQM11, EQM21, EQM31, EQM41); EQM1

#AIC e vies
>par(mfrow=c(2,2))
>plot(vetor_aic, vies11,main = "", xlab = "AIC ",
ylab="Vies beta1", ylim=c(-1, 1.5))
>plot(vetor_aic, vies21,main = "", xlab = "AIC ",
ylab="Vies beta2", ylim=c(-1, 1.5))
>plot(vetor_aic, vies31,main = "", xlab = "AIC ",
ylab="Vies beta3", ylim=c(-1, 1.5))
>plot(vetor_aic, vies41,main = "", xlab = "AIC ",
ylab="Vies beta4", ylim=c(-1, 1.5))

#ESTUDO DOS DADOS REAIS

rm(list=ls(all=TRUE))

#####
#Tamanho da amostra
#####
n=11

```

```
#####
# AJUSTE DO MODELO CHANTER
#####
>require(nlme)

>x<-dias2

#####
#Comprimento
#####
#chutes iniciais comprimento
>a1=13.9; b1=148.8; c1=0.047; d1=0.011
>regr_chantercomp<-nls(comprimentomedia~((a*b)/(a+(b-a)*exp((-c/d)*
(1-exp(-d*x))))), start=c(a=a1, b=b1, c=c1, d=d1), data=dadoscompri)
>(ajuste1 <- summary(regr_chantercomp))
>theta.g1 <- ajuste1$coefficients[,1]
>sd1 <- ajuste1$parameters[,2]
>f1<-summary(regr_chantercomp)
>plot(dias,comprimento, xlab="Tempo (dias)", ylab="Comprimento (cm)",
ylim=c(1,13),xlim=c(0,185))
>par(new=TRUE)
>plot(dias2,comprimentomedia, axes=FALSE, ann=FALSE, pch=13,
cex=2, col=2, ylim=c(1,13),xlim=c(0,185),xlab="Tempo (dias)",
ylab="comprimento (cm)")
>par(new=TRUE)
>plot(comprimentomedia~dias2, main="",xlab="Tempo (dias)",
ylab="Comprimento (cm)", ylim=c(1,13), xlim=c(0,185))
>at=seq(1,185, 0.5)
>b1=((f1$coefficients[1]*f1$coefficients[2])/(f1$coefficients[1]+
(f1$coefficients[2]-f1$coefficients[1])*exp((-f1$coefficients[3]/
f1$coefficients[4])*(1-exp(-f1$coefficients[4]*at))))
>lines(at,b1,col="red", lwd=2)

#####
#Diametro
#####
#chutes iniciais diametro
>a2=13.9; b2=170; c2=0.04; d2=0.011

>regr_chanterdiam<-nls(diametromedia~((a*b)/(a+(b-a)*exp((-c/d)*
(1-exp(-d*x))))), start=c(a=a2, b=b2, c=c2, d=d2), data=dadosdiam)
>(ajuste2<- summary(regr_chanterdiam))
>theta.g2 <- ajuste2$coefficients[,1]
>sd2 <- ajuste2$parameters[,2]
>f2<-summary(regr_chanterdiam)
>plot(dias,diametro,ylim=c(0,11),xlim=c(0,185),xlab="Tempo(dias)",
ylab="Diametro (cm)",)
```

```

>par(new=TRUE)
>plot(dias2,diametromedia, axes=FALSE, ann=FALSE,pch=24,col="blue",
ylim=c(0,11),xlim=c(0,185), xlab="Tempo (dias)", ylab="diametro (cm)",)
>par(new=TRUE)
>plot(diametromedia~dias2, xlab="Tempo (dias)", ylab="Diametro (cm)",
main="",ylim=c(0,11), xlim=c(0,185))
>at=seq(1,185, 0.5)
>b2=((f2$coefficients[1]*f2$coefficients[2])/(f2$coefficients[1]+
(f2$coefficients[2]-f2$coefficients[1])*exp((-f2$coefficients[3]/
f2$coefficients[4])*(1-exp(-f2$coefficients[4]*at)))))
>lines(at,b2, col="blue", lwd=2)

#####
# AJUSTE DO MODELO LOGISTICO
#####
>require(nlme)

>x<-dias2

#####
#Comprimento
#####

#chutes iniciais comprimento por meio da linearizacao do mod. logistico
>in_log<-function(dias2,comprimentomedia){
>A0=max(comprimentomedia)+1
>yl=log((A0/comprimentomedia)-1)
>modl=lm(yl~dias2)
>summary(modl)

>B0=as.numeric(-1*coef(modl)[1])
>k0=as.numeric(-1*coef(modl)[2])
>vinlog=c(A0,B0,k0)
>return(vinlog)
>}
>A1l=in_log(dias2,comprimentomedia)[1]; A1l
>B1l=in_log(dias2,comprimentomedia)[2]; B1l
>k1l=in_log(dias2,comprimentomedia)[3]; k1l

>regr_logcomp<-nls(comprimentomedia~((A)/(1+exp(-B-K*dias2))),
start=c(A=A1l, B=B1l, K=k1l ),
>data=dadoscompri)

>f3<-summary(regr_logcomp)
>theta.g3 <- f3$coefficients[,1]
>plot(dias,comprimento, xlab="Tempo (dias)", ylab="Comprimento
(cm)",ylim=c(1,13),xlim=c(0,185))

```

```

>par(new=TRUE)
>plot(dias2 ,comprimentomedia , axes=FALSE, ann=FALSE, pch=13, cex=2,
col=2,ylim=c(1,13),xlim=c(0,185),xlab="Tempo (dias)", ylab=
"Comprimento (cm)")
>par(new=TRUE)
>plot(comprimentomedia~dias2 , main="",xlab="Tempo (dias)",
ylab="Comprimento (cm)", ylim=c(1,13), xlim=c(0,185))
>at=seq(1,185, 0.5)
>b3=f3$coefficients[1]/(1+exp(-f3$coefficients[2]-
f3$coefficients[3]*at))
>lines(at,b3,col="red", lwd=2)

#####
#Diametro
#####

#chutes iniciais diametro por meio da linearizacao do mod. logistico
>in_log<-function(dias2,diametromedia){
>A0=max(diametromedia)+1
>yl=log((A0/diametromedia)-1)
>modl=lm(yl~dias2)
>summary(modl)

>B0=as.numeric(-1*coef(modl)[1])
>k0=as.numeric(-1*coef(modl)[2])
>vinlog=c(A0,B0,k0)
>return(vinlog)
>}

>A1=in_log(dias2,diametromedia)[1];A1
>B1=in_log(dias2,diametromedia)[2];B1
>k1=in_log(dias2,diametromedia)[3];k1

>regr_logdiam<-nls(diametromedia~((A)/(1+exp(-B-K*dias2))),
start=c(A=A1, >B=B1, K=k1 ),
>data=dadosdiam)

>f4<-summary(regr_logdiam)
>f4
>theta.g4 <- f4$coefficients[,1]
>plot(dias,diametro,ylim=c(0,11),xlim=c(0,185),xlab="Tempo (dias)",
ylab="Diametro (cm)",)
>par(new=TRUE)
>plot(dias2,diametromedia, axes=FALSE, ann=FALSE, pch=24,col="blue",
ylim=c(0,11),xlim=c(0,185), xlab="Tempo (dias)", ylab="Diametro (cm)",)
>par(new=TRUE)
>plot(diametromedia~dias2, xlab="Tempo (dias)", ylab="Diametro (cm)",
main="",ylim=c(0,11), xlim=c(0,185))

```



```

>b4=f4$coefficients[1]/(1+exp(-f4$coefficients[2]-f4$coefficients[3]
*at))
>lines(at,b4, col="blue", lwd=2)

#####
# AJUSTE DO MODELO GOMPERTZ
#####

>require(nlme)

>x<-dias2

#####
#Comprimento
#####

#chutes iniciais comprimento por meio da linearizacao do mod. gompertz
>alpha0 <- max(comprimentomedia) + 1
>z0 <- log(-log(comprimentomedia/alpha0))
>fit <- lm(z0~dias2)
>coef(fit)
# Valor inicial do parametro beta 1
>beta1p1=-coef(fit)[2]
>beta1p1
# Valor inicial do parametro beta 2
>beta2p1=coef(fit)[1]/beta1p1
>beta2p1

>regr_gompcomp <- nls(comprimentomedia ~ t*exp(-exp(-b1*(x-b2))),
start=list(t=alpha0, b1=beta1p1, b2=beta2p1))
>summary(regr_gompcomp)

>f5<-summary(regr_gompcomp)
>ajuste5 <- summary(regr_gompcomp)
>theta.g5 <- ajuste5$coefficients[,1]
>sd5 <- ajuste5$parameters[,2]
>plot(dias,comprimento, xlab="Tempo (dias)", ylab="Comprimento (cm)",
ylim=c(1,13),xlim=c(0,185))
>par(new=TRUE)
>plot(dias2,comprimentomedia, axes=FALSE, ann=FALSE, pch=13,
cex=2,col=2,ylim=c(1,13),xlim=c(0,185),xlab="Tempo (dias)",
ylab="Comprimento (cm)")
>par(new=TRUE)
>plot(comprimentomedia~dias2, main="",xlab="Tempo (dias)",
ylab="Comprimento (cm)", ylim=c(1,13), xlim=c(0,185))
>at=seq(1,185, 0.5)
>b5=f5$coefficients[1]*exp(-exp(-f5$coefficients[2]*

```

```

(at - f5$coefficients[3]))
>lines(at,b5,col="red", lwd=2)

#####
#Diametro
#####

>alpha01 <- max(diametromedia) + 1
>z01 <- log(-log(diametromedia/alpha01))
>fit1 <- lm(z01~dias2)
>coef(fit1)
# Valor inicial do parametro beta 1
>beta1p11=-coef(fit1)[2]
>beta1p11
# Valor inicial do parametro beta 2
>beta2p11=coef(fit1)[1]/beta1p11
>beta2p11

>regr_gompdiam <- nls(diametromedia ~ w*exp(-exp(-b11*(x-b21))),
  start=list(w=alpha01, b11=beta1p11, b21=beta2p11))
>summary(regr_gompdiam)

>f6 <- summary(regr_gompdiam)
>ajuste6 <- summary(regr_gompdiam)
>theta.g6 <- ajuste6$coefficients[,1]
>sd6 <- ajuste6$parameters[,2]
>plot(dias,diametro,ylim=c(0,11),xlim=c(0,185),xlab="Tempo (dias)",
  ylab="Diametro (cm)",)
>par(new=TRUE)
>plot(dias2,diametromedia, axes=FALSE, ann=FALSE, pch=24,col="blue",
  ylim=c(0,11),xlim=c(0,185), xlab="Tempo (dias)", ylab="Diametro (cm)",)
>par(new=TRUE)
>plot(diametromedia~dias2, xlab="Tempo (dias)", ylab="Diametro (cm)",
  main="",ylim=c(0,11), xlim=c(0,185))
>b6=f6$coefficients[1]*exp(-exp(-f6$coefficients[2]*(at -
  f6$coefficients[3])))
>lines(at,b6, col="blue", lwd=2)

```