

**Universidade de São Paulo
Escola Superior de Agricultura Luiz de Queiroz**

**Avaliação da qualidade de ajuste e predição de modelos não lineares:
uma aplicação em dados de crescimento de frutos de cacaueteiro**

Raquel Aline Oliveira Eloy

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

**Piracicaba
2018**

Raquel Aline Oliveira Eloy
Licenciada em Matemática

**Avaliação da qualidade de ajuste e predição de modelos não lineares:
uma aplicação em dados de crescimento de frutos de cacauero**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Profa. Dra. **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**

Eloy, Raquel Aline Oliveira

Avaliação da qualidade de ajuste e predição de modelos não lineares: uma aplicação em dados de crescimento de frutos de cacau / Raquel Aline Oliveira Eloy. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2018 .

48 p.

Dissertação (Mestrado) -- USP / Escola Superior de Agricultura "Luiz de Queiroz".

1.Modelo logístico 2.Modelo Gompertz 3.Modelo Von Bertalanffy 4.Qualidade de ajuste de predição 5.Cacau . I. Título.

DEDICATÓRIA

Minha Mãe, Raquel Oliveira, e a Záliston Melo, por todo carinho, amor e aprendizado, dedico.

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus, pelo dom da vida e por me sustentar nos momentos mais difíceis, aos meus pais, Geciro Soares (in memorian) e Raquel Souza, por tudo que fizeram por mim. Agradeço a minha família por ter me ajudado principalmente nesses dois anos fora de casa. Obrigada <3.

Ao departamento de Estatística da Universidade Estadual da Paraíba, em especial a Ana Patrícia Bastos Peixoto e Tiago Almeida de Oliveira, pelo incentivo e apoio.

Agradecer a minha orientadora e amiga Taciana Villela Savian, alguém que tem amor pelo que faz, mulher forte e amorosa e que no meio de tantas obrigações não esquece das coisas essenciais da vida <3.

Ao Departamento de Ciências Exatas - LCE, em especial aos professores Sônia Stefano, Roseli Leandro, Cristian Villegas e César Lima, obrigada por todo conhecimento, agradecer aos funcionários do departamento, em especial a Solange, por toda ajuda e formação.

Ao apoio financeiro da CAPES para o desenvolvimento e realização desse trabalho.

Os meus vizinhos amigos, que me acompanham e me incentivam desde a época da alfabetização até a festa de comemoração por ter sido aprovada no mestrado. Em especial, Antônio Lima e Família "alfabetização", Adriano Paulo e Família, Angélica Vasconcelos e Família, Maria José e Família, Laércia Cruz, Kathia Moreia, Juliana Santuza, Joana Erasmo e Família, Marilene Santos e Família, Vandenberg Santos e Família e Sônia Maria (in memorian). Maria Leite, pelo suporte em todas as horas.

Aos amigos e colegas que fiz nesses dois anos e que me ajudaram em algum momento dessa caminhada:

Abraão Taveira, por participar dessa jornada desde Campina Grande - PB, Altemir Braga, por ter me orientado e por estar sempre disponível, Andrezza Jardelino, por ter me recebido tão bem em Piracicaba, Douglas, Erasnilson e Rick Anderson pelos momentos de diversão e dúvidas esclarecidas, Tatiana Yamanouchi e Micheli Penso pela companhia, Alex Rocha (Alê) por escutar todas as reclamações, animar todas as conversas e aguentar meu sotaque (risos), a mana Patrícia Iwagaki, por todas as idas a padaria, risadas, sacodes e cumplicidade, te amo mana, Bruna Wruck, por sua praticidade em todos os momentos e ser a racional do grupo, Gustavo Thomas (raio de sol), pelo seu jeito de ser, Pollyane Vieira, por estar sempre incentivando a ser excelente, Aline Martinelli, amei as traças e dicas sobre ver o mundo com outros olhos, Cayán Portela pela diversão e bom humor, a

Mayara Carvalho, pelos abraços carinhosos e energia positiva e a Portya pela maneira de ver e aceitar os outros.

Em especial as pessoas com quem morei, Djair Durand, Fábio Prativiera, Julio Cezar e Leomir Sousa, fica aqui o meu obrigada a todos vocês, pela ajuda nos momentos difíceis e pelas comemorações, desejo a cada um de vocês o melhor. A república que me acolheu, devo muito a cada um de vocês, meu referencial de família em Pira <3.

Fábio, obrigada pelas aulas e pela paciência em estar sempre explicando os conteúdos. Deixa um registro pro Melado, mascote que trouxe alegria pra Rep.

Agradecer a quem está comigo há muito tempo e desde outras vidas, Záliston Melo, por entender minha ausência e por aguentar comigo mais uma fase.

Quando paro e olho pra trás fica difícil escrever o agradecimento, pela quantidade de pessoas que me ajudaram e me apoiaram durante toda a minha vida, seja com palavras, atitudes ou exemplo, então para todos que encontrei nesse caminho, deixo o meu sincero obrigada. “Sou uma parte de tudo aquilo que encontrei no meu caminho”.

SUMÁRIO

RESUMO	7
ABSTRACT	8
LISTA DE FIGURAS	9
LISTA DE TABELAS	10
1 INTRODUÇÃO	11
2 REVISÃO BIBLIOGRÁFICA	13
2.1 História do Cacau	13
2.2 Modelos de Regressão linear	15
2.3 Modelos de Regressão não linear	17
2.4 Método de Gauss-Newton	18
2.5 Modelos de crescimento	19
2.6 Intervalo de confiança do ponto de inflexão	21
2.7 Medidas estatísticas da qualidade do ajuste e predição	22
3 MATERIAIS E MÉTODOS	29
3.1 Banco de dados	29
3.2 Ajuste dos modelos	30
3.3 Técnicas de diagnóstico	31
4 RESULTADOS	33
5 CONCLUSÃO	41
REFERÊNCIAS	43
APÊNDICES	47

RESUMO

Avaliação da qualidade de ajuste e predição de modelos não lineares: uma aplicação em dados de crescimento de frutos de cacauero

Atualmente o Brasil é o quinto maior produtor de cacau no mundo e a sua produção está diretamente ligada ao ponto certo da colheita, colher frutos verdes ou verdoengos faz com que suas sementes tenham menor peso, em 1000 frutos maduros as amêndoas secas pesam em média 40 kg, colher frutos verdoengos, ou seja, frutos que estão parcialmente maduros, faz com que esse peso caia para 36 kg em média uma perda verificada de 10%, já quando os frutos estão verdes o peso passa a ser em média de 32 kg, possuindo uma perda de 20%, por isso é importante conhecer as fases de crescimento, que permite estabelecer formas adequadas de manejo, adubação e irrigação. Dentre as características biométricas do fruto do cacauero as que tem maior relevância econômica são o comprimento, o diâmetro e o volume. Uma forma de explicar relações de crescimento e produtividade de plantas, árvores, frutos ou animais é por meio da utilização de modelos de crescimento, pois possuem parâmetros com interpretação biológica. Os mais utilizados nestas áreas são os modelos: Logístico, Gompertz, Von Bertalanffy, Richards e Brody, sendo os dois últimos mais utilizados para descrever o crescimento animal. O objetivo do trabalho é avaliar a qualidade de ajuste e de predição dos modelos não lineares, Logístico, Gompertz e Von Bertalanffy, para medidas de comprimento e diâmetro do fruto do cacau, com a finalidade de prever o seu volume. Para prever o volume do fruto foi utilizado a fórmula de volume do esferoide prolato. Os critérios AIC e BIC foram utilizados para verificar qual dos modelos se ajusta melhor à essas medidas, já para verificar qual modelo se ajusta melhor na predição do volume do fruto, foram relacionadas as estatísticas: viés médio, índice de concordância, eficiência da modelagem e o desdobramento do quadrado médio do erro de predição.

Palavras-chave: Modelos não lineares; Viés médio; Índice de concordância; Eficiência da modelagem; Desdobramento do quadrado médio do erro de predição; Cacau

ABSTRACT

Evaluation of the quality of fit and prediction of nonlinear models: an application in growth data of cacao fruits

Currently Brazil is the fifth largest producer of cocoa in the world and its production is directly linked to the right point of harvest, harvesting green or green fruits makes their seeds have less weight, in 1000 mature fruits dry almonds weight on average 40 kg, harvested green fruits, that is, fruits that are partially ripe, causes this weight to fall to 36 kg on average a verified loss of 10%, when the fruits are green the weight becomes an average of 32 kg, with a loss of 20%, so it is important to know the growth phases, which allows to establish appropriate forms of management, fertilization and irrigation. Among the biometric characteristics of the cacao fruit, the most economically important are length, diameter and volume. One way to explain growth and productivity relationships of plants, trees, fruits or animals is through the use of growth models, since they have parameters that with biological interpretation are the most used in these areas: Logistics, Gompertz, Von Bertalanffy, Richards and Brody, the last two being most used to describe animal growth. The objective of this work is to evaluate the quality of fit and prediction of the non linear models, Logistic, Gompertz and Von Bertalanffy, for measures of length and diameter of the fruit of the cocoa, in order to predict its volume. To predict the volume of the fruit, the volume formula of the prolate spheroid was used. The AIC and BIC criteria were used to verify which model best fits these measures, and to verify which model best fits the fruit volume prediction, the statistics were related: mean bias, concordance index, modeling efficiency, and the unfolding of the mean square of the prediction error.

Keywords: Nonlinear models; Medium bias; Index of agreement; Modeling efficiency; Middle prediction error square scrolling; cocoa

LISTA DE FIGURAS

2.1 Fruto do cacau	14
2.2 Cacaueiro	14
2.3 Processo de polinização artificial	15
2.4 Modelos Logístico, Gompertz e Von Bertalanffy, com parâmetros $\beta_0 = 9$, $\beta_1 = 0, 2$, $\beta_2 = 0, 3$	20
4.1 Box-plot dos dados do comprimento, em centímetros, de cacau ao longo dos dias após a polinização.	33
4.2 Box-plot dos dados diâmetro, em centímetro, de cacau ao longo dos dias após a polinização.	34
4.3 Médias e ajuste do modelo Logístico, Gompertz e Von Bertalanffy dos dados do comprimento, em centímetro, de cacau em função dos dias.	34
4.4 Médias e ajuste dos modelos Logístico, Gompertz e Von Bertalanffy dos dados de diâmetro, em centímetro, de cacau em função dos dias.	35
4.5 Volume observado e estimado do cacau em cm^3 para os modelos (a) Lo- gístico, (b) Gompertz e (c) Von Bertalanffy em função do tempo após a polinização	38
4.6 Resíduos padronizados para a medida de comprimento do cacau para os modelos Logístico, Gompertz e Von Bertalanffy, respectivamente	39
4.7 Resíduos padronizados para a medida de diâmetro do cacau para os modelos Logístico, Gompertz e Von Bertalanffy, respectivamente	40
4.8 Gráfico Normal de probabilidade para a medida de comprimento do cacau para os modelos Logístico, Gompertz e Von Bertalanffy, respectivamente. .	40
4.9 Gráfico Normal de probabilidade para a medida de diâmetro do cacau para os modelos Logístico, Gompertz e Von Bertalanffy, respectivamente.	40

LISTA DE TABELAS

2.1	Modelos de regressão não linear para descrever curvas de crescimento. . . .	20
3.1	Idade, número de frutos avaliados, médias das medidas de comprimento, diâmetro e volume do fruto do Cacaueiro.	29
4.1	Estimativas dos parâmetros (erro padrão) e avaliadores da qualidade de ajuste do modelo Logístico, na descrição de comprimento e diâmetro dos frutos do cacaueiro.	35
4.2	Estimativas dos parâmetros (erro padrão) e avaliadores da qualidade de ajuste do modelo Gompertz, na descrição de comprimento e diâmetro dos frutos do cacaueiro.	36
4.3	Estimativas dos parâmetros (erro padrão) e avaliadores da qualidade de ajuste do modelo Von Bertalanffy, na descrição de comprimento e diâmetro dos frutos do cacaueiro.	36
4.4	Coordenada da abscissa do ponto de inflexão estimadas com os modelos de crescimento Logístico, Gompertz e Von Bertalanffy.	37
4.5	Medidas estatísticas da qualidade de predição do volume do fruto do cacaueiro.	38
4.6	Medida de não linearidade para as medidas de comprimento e diâmetro . .	39

1 INTRODUÇÃO

Segundo MARTINS *et al.* (2011) a colheita do cacau deve ser feita de forma planejada, pois colher frutos que não atingiram o estágio ideal de maturação, acarreta em sementes com menor peso e teor de açúcar, o que prejudica a qualidade do produto final. Para estudar o crescimento de espécies, sejam animais ou vegetais é de extrema importância explicar relações de crescimento e produtividade de plantas, frutos, animais, etc. O estudo de modelos de crescimento, que são não lineares em relação aos seus parâmetros, é uma das alternativas mais empregadas para estudar o crescimento de espécies, Os modelos de crescimento mais comumente utilizados nestas áreas são os modelos: Logístico (NELDER, 1961), Gompertz (LAIRD *et al.*, 1965) e Von Bertalanffy (VON BERTALANFFY, 1957). No caso do cacau, desde o final da década de 80 o setor cacauero vem sofrendo uma grave crise ocasionada pela doença denominada vassoura-de-bruxa, para contornar essa situação pesquisas estão sendo desenvolvidas no que diz respeito a novos cultivares.

Quando vários modelos são ajustados a um conjunto de dados é importante, e geralmente de interesse do estudo, saber qual dos modelos possui o melhor ajuste e capacidade de predição da variável resposta e diversos critérios são empregados com essa finalidade. O objetivo deste trabalho consiste em ajustar os modelos Logístico, Gompertz e Von Bertalanffy aos dados de comprimento e diâmetro do fruto do cacauero, com a finalidade de prever o volume do fruto do cacau utilizando um modelo matemático. No experimento original o volume do fruto foi tomado usando a teoria de Arquimedes, que consiste em mergulhar o fruto em um recipiente e a água deslocada corresponde ao volume do fruto. Esse estudo foi realizado com métodos destrutivos, o que acarreta na modificação do sistema fisiológico que alimenta os frutos restantes. A proposta deste trabalho foi estimar o volume do fruto do cacau usando um modelo matemático, por não ser um método destrutivo, preserva o fruto do cacauero e faz com que o fruto seja colhido no estágio certo de maturação. Para avaliar a qualidade de ajuste dos modelos não lineares foram utilizados os critérios de Akaike (AIC), critério bayesiano (BIC), o desvio padrão residual e a medida de curvatura de Bates e Watts, já no que se refere a predição do volume por meio do modelo matemático, foram utilizados o viés médio, o índice de concordância, a eficiência da modelagem bem como o desdobramento do quadrado médio do erro de predição, sendo esse último proposto por KOBAYASHI e SALAM (2000).

Nesta primeira seção apresentamos toda a estrutura do trabalho. Na 2ª seção apresentamos uma breve história sobre o cacau, definimos modelos de regressão linear e não linear, descrevemos o método de Gauss-Newton, adicionalmente apresentamos os modelos de crescimento e o intervalo de confiança do ponto de inflexão desses modelos. Apresenta-se, ainda, medidas estatísticas da qualidade do ajuste de predição. Na seção 3 apresentamos o banco de dados utilizados, o ajuste dos modelos e técnicas de diagnóstico. Na seção 4 apresentamos a aplicação da metodologia discutida, bem como todos os

resultados. Por fim, na seção 5 conclusões com base nos resultados obtidos. No Apêndice A encontram-se cálculos especificados de forma detalhada e no Apêndice B o código computacional que foi realizado no desenvolvimento deste trabalho.

2 REVISÃO BIBLIOGRÁFICA

2.1 História do Cacau

Carl Nilsson Linnaeus, famoso botânico sueco considerado o pai da taxonomia moderna, batizou o cacau com o nome de *Théobroma*, que significa alimento dos deuses, pois a história do cacau é uma mistura de lenda mexicana e mitologia, onde Quetzalcoatl, Deus asteca da lua, roubou a planta do país dos filhos do sol para dá-las aos homens, oferecendo-lhes assim alegrias desconhecidas. O cacau foi tão valorizado que era usado como moeda pela civilização asteca (BATISTA, 2009).

Suas sementes, levadas para outras regiões e continentes, formaram grandes plantações que, hoje, representam importante fonte de trabalho e renda para milhões de pessoas. As primeiras plantações africanas foram feitas por volta de 1855, nas ilhas de São Tomé e Príncipe, colônias portuguesas ao largo da costa ocidental africana (MARTINS *et al.*, 2011) .

O Brasil passou por três grandes ciclos econômicos, o da cana-de-açúcar, entre os séculos XVI e XVIII no período colonial, o do cacau, de 1889 à 1930, e do café no período entre 1800 e 1930. Cada um deles foi responsável pela riqueza e transformação de um determinado período, sendo que seu apogeu e sua decadência influenciaram na organização do espaço brasileiro presenciado atualmente. O ciclo cacauzeiro foi de extrema relevância ao país no período da República Velha (1889-1930), pois devido ao desenvolvimento da indústria mundial de chocolate, aumentou a demanda pela sua matéria-prima, o cacau. Com o fruto do cacauzeiro prepara-se o chocolate, comum em bebidas, bombons, bolos, doces, sorvetes, além da extração da manteiga de cacau e produção de cosméticos, estudos recentes o colocam como muito benéfico à saúde (BATISTA e VIERO, 2014).

Em relação ao comércio internacional, o Brasil participa exportando cacau em amêndoas desde 1961; mas, a partir de 1992, passou a importar um total de 336 mil toneladas até o ano de 2002. O cacau é uma cultura muito importante na formação de receita por hectare e na contribuição no Valor Bruto da Produção Agrícola (VBPA) na região dos Tabuleiros Costeiros da Bahia (TC/BA). No ano 2002, nas microrregiões da Bahia onde a cultura é praticada, o cacau chegou a contribuir com 66% do valor total das culturas permanentes em 2002 (CUENCA e NAZÁRIO, 2004).

A Conab (Companhia Nacional de Abastecimento), registrou em 2016 que mundialmente a África é responsável por 74% da produção do cacau, liderada pela Costa do Marfim, Camarões, Gana e Nigéria, ficando assim o Brasil como o quinto maior produtor de cacau no mundo, 95% da sua produção está concentrada no sul da Bahia, 3,5% no Espírito Santo e 1,5% no estado da Amazônia. A produção está diretamente ligada ao ponto certo da colheita, por isso é importante conhecer as fases de crescimento, que permite estabelecer formas adequadas de manejo, adubação e irrigação.

O cacauzeiro é uma planta da família das *Sterculiáceae*, no Brasil oficialmente

teve início em 1679, com a autorização da Carta Régia. Várias tentativas foram feitas no estado do Pará, mas o cultivo não se estabeleceu naquela época, devido a pobreza dos solos da região. No século XVIII um agricultor da Bahia recebeu algumas sementes de um colonizador do Pará e introduziu o cultivo no estado. O clima quente e úmido da região fez com que rapidamente o sul do estado se transformasse no principal produtor de cacau do país (MARTINS *et al.*, 2011).

Atinge uma altura entre 5 e 10 metros na maturidade, torna-se adulta aos 10 anos, e pode, em circunstâncias excepcionais, produzir até aos 50 anos de idade. O tamanho da fruta madura varia entre 15 a 30 centímetros de comprimento por 8 a 13 de largura. Cada fruto contém entre 20 e 40 sementes envoltas numa polpa macia de cor castanha esbranquiçada. O cacauzeiro necessita de calor e umidade, temperatura entre 24°C e 28°C, e no mínimo 1.500 mm de chuva, distribuídos pelo ano inteiro, o mais regularmente possível (HERMÉ, 2007).

O cacau produz o ano inteiro, no entanto duas épocas ganham destaque. Uma principal (safra), que pela produção ser maior permite um frutos de maior qualidade, que vai de maio a setembro e uma outra secundária (temporão), cuja produção é menor que vai de outubro a março. O solo deve ser profundo e bem drenado, os mais indicados são solos vermelho-escuro, o prodizólico vermelho-amarelado e solos aluviais de boa fertilidade natural.



Figura 2.1. Fruto do cacau

Fonte: Revista Globo Rural



Figura 2.2. Cacaueiro

Fonte: Cacao (Cocoa Powder) Facts, Uses, Health Benefits, Side Effects

A reprodução do cacauero é realizada por polinização ou com sementes, em viveiros. No viveiro a muda é transplantada para o lugar definitivo no sexto mês, plantadas a quatro metros umas das outras. A colheita do cacau é realizada fruto por fruto, utilizando o podão, tendo cuidado para não estragar os brotos e as flores de cacau (MARTINS *et al.*, 2011). O principal inimigo desse cultivo é a vassoura-de-bruxa, praga que surgiu



Figura 2.3. Processo de polinização artificial

Fonte: CEPLAC

na bacia amazônica e é uma das doenças fitopatológicas mais graves do Brasil. Plantar mudas modificadas geneticamente vem se mostrando a melhor maneira de lidar com essa praga, estas espécies são resistentes, mas não imunes e com o passar do tempo podem se tornar vulneráveis (EFRAIM, 2009).

No Brasil, os primeiros trabalhos de melhoramento conduzidos com cacau incluíram a seleção de plantas em cultivos comerciais e a avaliação de variedades clonais, que resultaram nas séries SIC, SIAL e EEG produzidas, respectivamente, pelo Instituto do Cacau, Instituto Agrônomo do Leste e Estação experimental de Goitacazes, no entanto não chegaram a ter representantes distribuídos para o plantio comercial (PIRES *et al.*, 2003).

2.2 Modelos de Regressão linear

O termo regressão foi utilizado pela primeira vez por Sir Francis Galton em 1889, que 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. Este acontecimento foi denominado de regressão para a estatura média, atualmente o termo regressão para estatura média não é mais utilizado, permanecendo apenas o termo regressão.

Um modelo de regressão linear descreve uma relação entre a variável dependente, também denominada variável resposta (Y_i) e uma ou mais variáveis independentes ou explicativas, X_1, X_2, \dots, X_k . Dada uma amostra de n observações, temos, $X_{1i}, X_{2i}, \dots, X_{ki}, Y_i$,

$i=1,2,\dots,n$. Dessa forma o modelo estatístico que descreve a relação entre as variáveis é dado por:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \epsilon_i$$

em que: $\beta_0, \beta_1, \dots, \beta_k$ são os parâmetros do modelo, que quantificam a influência que uma variável explicativa tem sobre a variável resposta, Y_i é a i -ésima observação da variável resposta; X_{ji} é a i -ésima observação da j -ésima variável com $j=1,\dots,k$ e ϵ_i é o erro experimental associado a i -ésima observação.

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 não dependem destes. Conforme apresentado no trabalho de FIGUEIREDO FILHO *et al.* (2011) alguns pressupostos precisam ser satisfeitos:

- (1) a relação entre a variável dependente (Y_i) e as variáveis independentes (X_{ji}) deve ser linear.
- (2) as variáveis devem ser medidas adequadamente, ou seja, assume-se que não há erro sistemático de mensuração.
- (3) as variáveis independentes (X_{ji}) são não correlacionadas – hipótese de ausência de multicolinearidade entre as variáveis explicativas
- (4) homocedasticidade, ou seja, a variância do termo de erro é constante para os diferentes valores da variável independente.
- (5) os erros são independentes com média nula.
- (6) assume-se que os erros ϵ_i tem uma distribuição normal.

Existem vários métodos para a estimação dos parâmetros de um modelo, o método dos mínimos quadrados (MMQ) desenvolvido por Carl Friedrich Gauss em 1809, 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 SQR torna-se mínima. Considerando o modelo ajustado, dado por:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki},$$

temos que a soma de quadrado dos resíduos pode ser expressa como:

$$SQR = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki})]^2.$$

Cada um dos estimadores $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ é uma função linear nos parâmetros, se os valores de Y_i são normalmente distribuídos e são independentes entre si, os estimadores terão distribuição normal, com desvios padrões facilmente obtidos. Considere $\hat{\epsilon} = (\mathbf{y} - \mathbf{X}\hat{\eta})$,

logo SQR pode ser reescrita na forma matricial:

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

Diferenciando em relação a $\boldsymbol{\beta}$, temos:

$$\frac{\partial \hat{\epsilon}' \hat{\epsilon}}{\partial \boldsymbol{\beta}} = 0 - 2\mathbf{X}' \mathbf{y} + 2\mathbf{X}' \mathbf{X} \boldsymbol{\eta}.$$

Igualando a zero obtemos o sistema de equações normais:

$$\mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}' \mathbf{y}.$$

Portanto, $\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$ é o estimador de mínimos quadrados para o modelo de regressão linear. Existem situações em que o modelo linear não é o mais apropriado, nestes casos modelos de regressão não linear podem ser utilizados, assunto que será abordado na próxima seção.

2.3 Modelos de Regressão não linear

Os modelos de regressão não lineares, segundo SMYTH (2006), podem ser escritos da seguinte forma:

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}_i; \boldsymbol{\beta}) + \boldsymbol{\epsilon} \quad (2.1)$$

em que: \mathbf{Y} é um vetor ($n \times 1$) de observações, $\mathbf{f}(\mathbf{X}_i; \boldsymbol{\beta})$ é um vetor ($n \times 1$) de funções não lineares em $\boldsymbol{\beta}$, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)'$ é um vetor com $p+1$ parâmetros e $\boldsymbol{\epsilon}$ é um vetor ($n \times 1$) dos erros aleatórios, assumindo geralmente como não correlacionados com média zero e variância constante.

Segundo MAZUCHELI e ACHCAR (2008), um modelo de regressão não linear tem pelo menos um de seus parâmetros apresentando não linearidade, ou seja, quando ao menos uma das derivadas parciais da função, em relação aos parâmetros, depende de pelo menos um dos parâmetros do modelo, este é dito não linear.

Na modelagem de dados de crescimento, deseja-se obter informações sobre a interpretação prática dos parâmetros e para modelar esse tipo de comportamento a literatura apresenta vários modelos de crescimento. Os apresentados nesse trabalho são os modelos Logístico, Gompertz e Von Bertalanffy.

Na obtenção das estimativas dos parâmetros de um modelo de regressão não linear, assim como nos modelos de regressão linear, podemos utilizar o método dos mínimos quadrados ordinários que consiste em minimizar a soma dos quadrados residuais.

Matricialmente, a soma de quadrados dos resíduos, $SQR(\boldsymbol{\beta})$, do modelo (2.1) pode ser escrita como:

$$SQR(\boldsymbol{\beta}) = \sum_{i=1}^n [Y_i - f(X_i; \boldsymbol{\beta})]^2,$$

Na expressão de $SQR(\boldsymbol{\beta})$ para encontrar o estimador de mínimos quadrados, é necessário diferenciar $SQR(\boldsymbol{\beta})$ com relação a cada um dos parâmetros e igualar essa equação a zero, ou seja,

$$\frac{\partial SQR(\boldsymbol{\beta})}{\partial \beta_r} \Big|_{\hat{\boldsymbol{\beta}}} = 0 \quad (r = 1, 2, 3, \dots, p)$$

$$\sum_{i=1}^n [Y_i - f(X_i; \boldsymbol{\beta})] \frac{\partial f(X_i; \boldsymbol{\beta})}{\partial \beta_r} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0 \quad (r = 1, 2, 3, \dots, p)$$

Assim, obtém-se um sistema de p equações normais não lineares. Geralmente o sistema de equações normais (SEN), no caso de modelos de regressão não linear não possui solução analítica, assim, a obtenção das estimativas dos parâmetros é realizada utilizando métodos numéricos iterativos.

Na literatura são apresentados vários métodos iterativos, são exemplos, o método de Newton-Raphson, que consiste em uma aproximação da raiz da equação pela derivada da função, necessitando de um critério de parada e uma boa estimativa de valor inicial; o algoritmo EM, que segundo DEMPSTER *et al.* (1977) é utilizado para calcular o estimador de máxima verossimilhança de forma iterativa, principalmente no caso em que os dados são incompletos, é descrito resumidamente em dois passos, o passo “E”, que calcula o valor esperado do logaritmo da função de verossimilhança; e o passo “M”, que encontra seu máximo, sendo repetido até atingir a convergência. Há também o método de Gauss-Newton (HARTLEY, 1961a) que será descrito em mais detalhes na seção que se segue.

2.4 Método de Gauss-Newton

O método de Gauss-Newton (HARTLEY, 1961b), também conhecido como método da linearização, usa uma expansão em série de Taylor de primeira ordem para aproximar o modelo de regressão não linear com termos lineares e, então, aplicar o método dos mínimos quadrados ordinários para estimar os parâmetros.

Suponha que $\boldsymbol{\beta}$ é uma aproximação da estimativa de mínimos quadrados $\boldsymbol{\beta}^{(0)}$ de um modelo não linear, tomando $\boldsymbol{\beta}$ próximo de $\boldsymbol{\beta}^{(0)}$, tem-se que a expressão da série de Taylor de primeira ordem é

$$\mathbf{f}(\mathbf{x}, \boldsymbol{\beta}) \approx \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}^{(0)}) + \mathbf{F}(\boldsymbol{\beta}^{(0)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}),$$

em que $\mathbf{F}(\boldsymbol{\beta}^{(0)})$ é a matriz de primeiras derivadas de $\boldsymbol{\beta}$. Definindo $\mathbf{s}(\boldsymbol{\beta})$ como sendo o

vetor de resíduos, obtemos

$$\mathbf{s}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}) \approx \mathbf{s}(\boldsymbol{\beta}^{(0)}) - \mathbf{F}^{(0)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}),$$

onde $\mathbf{F}^{(0)} = \mathbf{F}(\boldsymbol{\beta}^{(0)})$. Substituindo $\mathbf{s}'(\boldsymbol{\beta})\mathbf{s}(\boldsymbol{\beta})$ em $SQR(\boldsymbol{\beta})$, temos

$$\begin{aligned} SQR(\boldsymbol{\beta}) &\approx \mathbf{s}'(\boldsymbol{\beta}^{(0)})\mathbf{s}(\boldsymbol{\beta}^{(0)}) - 2\mathbf{s}'(\boldsymbol{\beta}^{(0)})\mathbf{F}^{(0)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) + (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})'\mathbf{F}'^{(0)}\mathbf{F}^{(0)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}). \\ &= \mathbf{s}(\boldsymbol{\beta}^{(0)})'\mathbf{s}(\boldsymbol{\beta}^{(0)}) - 2\mathbf{s}(\boldsymbol{\beta}^{(0)})\mathbf{F}(\boldsymbol{\beta}^{(0)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) + (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})'\mathbf{F}(\boldsymbol{\beta}^{(0)})'\mathbf{F}(\boldsymbol{\beta}^{(0)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}). \end{aligned}$$

Em que essa expressão pode ser vista como expressão matricial para a soma de quadrados de uma regressão linear simples com $(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) = \boldsymbol{\beta}$, $\mathbf{F}(\boldsymbol{\beta}^{(0)}) = \mathbf{X}$ e $\mathbf{s}(\boldsymbol{\beta}^{(0)}) = \mathbf{y}$

Portanto, $SQR(\boldsymbol{\beta})$ será minimizada quando

$$\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)} = [\mathbf{F}'^{(0)}\mathbf{F}^{(0)}]^{-1}\mathbf{F}'^{(0)}\mathbf{s}(\boldsymbol{\beta}^{(0)}).$$

Assim, devido à aproximação $\boldsymbol{\beta}^{(0)}$, a próxima aproximação é dada por

$$\boldsymbol{\beta}^{(0+1)} = \boldsymbol{\beta}^0 + [\mathbf{F}'^{(0)}\mathbf{F}^{(0)}]^{-1}\mathbf{F}'^{(0)}\mathbf{s}(\boldsymbol{\beta}^{(0)}), \quad (2.2)$$

os valores iniciais podem ser obtidos de estudos anteriores, conhecimento teórico ou por uma grade de valores que minimize a soma de quadrados dos resíduos, repetindo até que o vetor de estimativas não se altere mais dentro de uma precisão pré estipulada.

2.5 Modelos de crescimento

O termo curva de crescimento remete a imagem de uma curva sigmóide, em formato de “S”, que descreve uma série de medidas de tamanho ao longo do tempo. Podem ser caracterizados de três formas: Estática, se uma ou várias medidas são observadas, apenas uma vez, para um grupo de indivíduos. Examinar esses dados fornece pouca informação sobre os padrões de crescimento e desenvolvimento. Seccional-cruzada, onde cada indivíduo é medido apenas uma vez, essas medições são feitas em outras idades (estágios) com outros indivíduos. A qualidade da informação vai depender da representação de cada indivíduo na população. Por fim, Longitudinal, inclui todas as informações disponíveis na seccional-cruzada e estática. A confiança na interpretação é aumentada uma vez que uma ou mais variáveis são medidas mais de uma vez em cada uma das unidades amostrais (FITZHUGH, 1976).

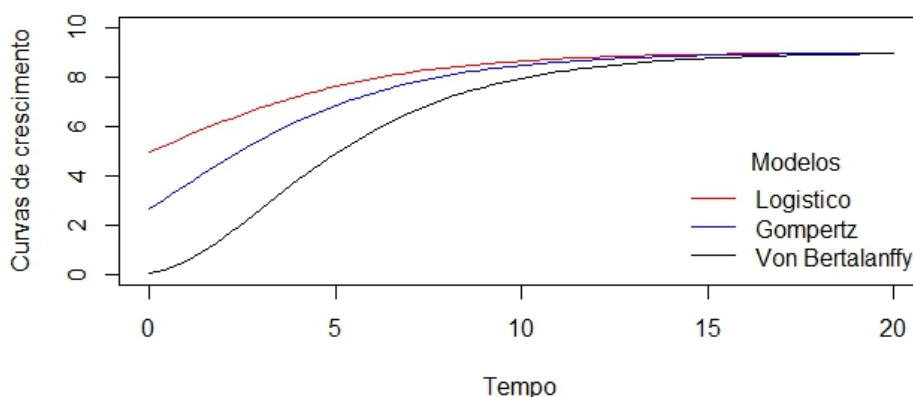
Os modelos de regressão, lineares e não lineares, são utilizados em diversas áreas do conhecimento, como na agronomia, zootecnia, fitotecnia, entre outras. Dentre os modelos não lineares, podemos destacar os modelos de crescimento que são bastante utilizados em estudos vegetal e animal. A Tabela 2.1 mostra os modelos de crescimento mais utilizados.

Tabela 2.1. Modelos de regressão não linear para descrever curvas de crescimento.

Modelos	Modelos Estatísticos	Espaço Paramétrico
Logístico	$Y_i = \frac{\beta_0}{[1 + \exp(-\beta_1 - \beta_2 X_i)]} + \epsilon_i.$	$\beta_0 > 0, \beta_1 \in \mathbb{R}, \beta_2 > 0$
Gompertz	$Y_i = \beta_0 \exp[-\exp(\beta_1 - \beta_2 X_i)] + \epsilon_i.$	$\beta_0 > 0, \beta_1 \in \mathbb{R}, \beta_2 > 0$
Von Bertalanffy	$Y_i = \beta_0 [1 - \exp(-\beta_1 - \beta_2 X_i)]^3 + \epsilon_i.$	$\beta_0 > 0, \beta_1 \in \mathbb{R}, \beta_2 > 0$

Os estudos de curvas de crescimento são importantes uma vez que podem ser realizados em situações em que a informação contida numa sequência de pontos “tamanho-tempo”, do crescimento de uma espécie seja reduzida e sintetizada num conjunto de parâmetros que tenham interpretação prática e preditiva (PRADO *et al.*, 2013).

MARTINS (2013) estudou a qualidade de ajuste dos modelos não lineares Logístico e Gompertz, para o crescimento de frutos de pereira asiáticas, utilizando o comprimento dos frutos e considerando independência e homogeneidade de variância dos erros. TERRA *et al.* (2010) utilizaram os modelos Logístico e Gompertz para estudar o crescimento da Tamareira-Anã (*Phoenix roebelenii O'brien*) e avaliou o ponto de maturidade fisiológica das sementes contidas no interior dos diásporos. PRADO *et al.* (2013) utilizaram os modelos não lineares de crescimento, Gompertz e Logístico, para realizar estudos no fruto de coqueiro da variedade anã verde. A Figura 2.4 mostra o comportamento dos modelos Logístico, Gompertz e Von Bertalanffy.

**Figura 2.4.** Modelos Logístico, Gompertz e Von Bertalanffy, com parâmetros $\beta_0 = 9$, $\beta_1 = 0, 2$, $\beta_2 = 0, 3$

Como os modelos Logístico, Gompertz e Von Bertalanffy possuem parâmetros de fácil interpretação são utilizados frequentemente na descrição do crescimento de frutos. O modelo Logístico foi proposto por Verhulst em 1838 para expressar a lei do crescimento

de populações humanas, no seu primeiro trecho aproxima-se a uma função exponencial, possuindo uma curva simétrica em relação ao ponto de inflexão dado por $I = \left(-\frac{\beta_1}{\beta_2}, \frac{\beta_0}{2}\right)$ e assíntota igual ao parâmetro β_0 . A relação $\left(\frac{-\beta_1}{\beta_2}\right) = t$, indica que em t dias ocorreu a velocidade máxima do crescimento por dia e após esse dia há uma desaceleração no crescimento do fruto. O β_2 é a taxa de expansão ou velocidade com que o fruto atinge seu valor assintótico.

O modelo Gompertz, criado por Benjamin Gompertz em 1825 para estudar as leis de natalidade e mortalidade humanas, é uma curva assimétrica em relação ao seu ponto de inflexão, ou seja, tem um desenvolvimento mais lento quando alcança esse ponto. Tem comportamento similar com a curva Logística, mas seu ponto de inflexão ocorre em, $I = \left(\frac{\beta_1}{\beta_2}, \frac{\beta_0}{e}\right)$, onde e é base dos logaritmos neperianos, a taxa de crescimento é máxima antes da metade do crescimento máximo.

Segundo BRITO *et al.* (2007), o Modelo Von Bertalanffy (1957) veio de estudos das relações alométricas dos organismos, sendo aplicado comumente em estudo de crescimento de peixes e possui ponto de inflexão, $I = \left(\frac{\ln 3 - \beta_1}{\beta_2}, \frac{8\beta_0}{27}\right)$. Neste modelo, o parâmetro β_0 representa o valor assintótico, β_1 não possui interpretação biológica bem definida e o β_2 é a taxa de maturação. Quanto maior o valor de β_2 , maior a velocidade de crescimento.

2.6 Intervalo de confiança do ponto de inflexão

O ponto de inflexão dos modelos não lineares é uma informação importante para o pesquisador, pois, aumentamos a inferência à respeito da produção ao longo do tempo, determinando assim, quando o crescimento do fruto está mais acelerado. Calculando a segunda derivada, em relação a cada parâmetro, e igualando a zero temos o ponto de inflexão. MISCHAN e DE PINHO (2014) propõem uma forma de estimar a variância da estimativa da abscissa do ponto de inflexão $I = (X_{pi}, Y_{pi})$, dos modelos Logístico, Gompertz e Von Bertalanffy. Tendo posse da abscissa do ponto de inflexão, derivando e elevando ao quadrado, a estimativa da variância ($S^2(X_{pi})$) pode ser escrita em termos das estimativas das variâncias e covariâncias das estimativas dos parâmetros.

Para o modelo Logístico, temos, como abscissa do ponto de inflexão a expressão, $X_{pi} = \frac{-\beta_1}{\beta_2}$, já para o modelo Von Bertalanffy, temos, como abscissa do ponto de inflexão (X_{pi}) a expressão, $X_{pi} = \frac{\ln 3 - \beta_1}{\beta_2}$ e para o modelo Gompertz, temos, como abscissa do ponto de inflexão (X_{pi}) a expressão, $X_{pi} = \frac{\beta_1}{\beta_2}$.

O estimador da variância da abscissa do ponto de inflexão é dada por:

$$S^2(X_{pi}) = \frac{1}{\beta_2^2} [f \hat{\beta}_1 \hat{\beta}_1 + X_{pi}^2 f \hat{\beta}_2 \hat{\beta}_2 - 2X_{pi} f \hat{\beta}_1 \hat{\beta}_2] \quad (2.3)$$

em que β_1 e β_2 são os parâmetros estimados do modelo e $f\hat{\beta}_1\hat{\beta}_1, f\hat{\beta}_2\hat{\beta}_2$ são as estimativas das variâncias das estimativas dos parâmetros β_1 e β_2 , respectivamente e $f\hat{\beta}_1\hat{\beta}_2$ é a estimativa da covariância entre as estimativas dos parâmetros β_1 e β_2 .

Para calcular o intervalo de confiança da abcissa do ponto de inflexão usaremos a expressão (PINHO *et al.*, 2014), dada por:

$$IC_{(X_{pi};95\%)} = X_{pi} \pm S(I).t_{(0,25;GL_{Erro})}$$

Onde X_{pi} é a abcissa do ponto de inflexão, $t_{(0,25;GL_{Erro})}$ é o quantil da distribuição t de Student com $\alpha = 5\%$ e GL_{Erro} graus de liberdade e $S(I)$ é a raiz quadrada de $S^2(I)$.

2.7 Medidas estatísticas da qualidade do ajuste e predição

Comumente, os trabalhos que avaliam a qualidade de ajuste de modelos de regressão não linear utilizam, para alcançar seus objetivos, critérios como o de informação de Akaike - AIC (AKAIKE, 1974) e o critério de informação Bayesiano - BIC (SCHWARZ *et al.*, 1978). Esses critérios são baseados no logaritmo da função de verossimilhança do modelo $l(\boldsymbol{\theta})$, no número de parâmetros do modelo (p) e no número de observações (n). As expressões dos critérios AIC e BIC são dadas, respectivamente, por:

$$AIC = -2l(\boldsymbol{\theta}) + 2p,$$

$$BIC = -2l(\boldsymbol{\theta}) + p \log(n).$$

Com o objetivo de investigar a não linearidade de um modelo, BATES e WATTS (1980) propuseram uma medida de curvatura que serve pra indicar o quão distante um modelo não linear está de um modelo linear, sendo decomposta em duas componentes, a não linearidade intrínseca, associada à curvatura do espaço de estimação e a não linearidade devido ao efeito de parâmetros. O modelo não linear estará mais próximo do linear quanto menor for o valor das duas componentes de não linearidade. Para isso os autores utilizaram a teoria de modelos lineares, um intervalo de confiança $1 - \alpha$ para $\boldsymbol{\theta}$ é obtido obedecendo a desigualdade

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{X}' X (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq ps^2 F_{p,n-p}^\alpha$$

em que \mathbf{X} é a matriz do modelo e $\boldsymbol{\theta}$ é o vetor de parâmetros. Estipularam a quantidade $\rho = s\sqrt{p}$ como raio padrão da esfera cujos valores internos $\boldsymbol{\theta}$ pertencem ao intervalo de confiança $1 - \alpha$.

A validade da aproximação linear depende da magnitude do segundo termo da expressão de série de Taylor relativo ao primeiro termo, dividindo assim a componente em dois termos, essa decomposição pode ser obtida definindo-se uma matriz de projeção ($\hat{\mathbf{P}}$), dada por:

$$\hat{\mathbf{P}} = \hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}', \quad (2.4)$$

dessa maneira, as duas componentes são dadas por:

$$\hat{\mathbf{H}} = \hat{\mathbf{H}}^T + \hat{\mathbf{H}}^N \quad (2.5)$$

em que os índices sobrescritos T e N indicam as componentes tangencial e normal ao plano definido pela aproximação linear. Definiram assim, essas medidas comparando o termo quadrático com o linear da seguinte forma:

$$K^T = \frac{\|(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{H}^T (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\|}{\|\mathbf{F}((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}))\|^2} \text{ e } K^N = \frac{\|(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{H}^N (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\|}{\|\mathbf{F}((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}))\|^2},$$

em que K^T e K^N representam as curvaturas dos círculos que melhor aproximam do formato da superfície esperada em um modelo $\eta(\hat{\boldsymbol{\theta}})$. Esses círculos possuem raio $R = \frac{1}{K}$ e, para obter uma medida invariante à escala da variável resposta, adotaram padronizar essas mediadas pelo raio padrão ρ . Então,

$$\Gamma^T = K^T \rho \text{ e } \Gamma^N = K^N \rho$$

são medidas adotadas para julgar a não linearidade das componentes do modelo em que Γ^T está associada ao efeito de parâmetro e Γ^N é a curvatura intrínseca do modelo. Maiores detalhes ver BEALE (1960).

Outra medida utilizada para avaliar a qualidade de ajuste de um modelo é o desvio padrão residual (RSD), dado por:

$$RSD = \sqrt{\frac{QME}{n - p}}$$

em que, QME é o quadrado médio do erro, n o número de observações e p o número de parâmetros do modelo. O conceito do erro padrão da estimativa ou desvio padrão residual, é equivalente ao do desvio padrão que mede a variabilidade dos valores da amostra ao redor da média aritmética desses valores, sendo assim, quanto mais próximo de zero melhor.

Além da qualidade de ajuste de um modelo também é importante avaliar a qualidade de predição das variáveis respostas e várias medidas estatísticas são utilizadas para avaliar essa qualidade de predição. Dentre elas daremos destaque para quatro medidas: O viés médio (MBE), a raiz do erro quadrático médio (RMSD), o índice de concordância (IA) e a eficiência de modelagem (ME).

Nas expressões (2.6), (2.7), (2.8), (2.9), (2.10) e (2.11), apresentadas a seguir, Y_i são os valores observados da variável resposta, \hat{Y}_i são os valores preditos pelo modelo, \bar{Y} é a média dos valores observados e $\bar{\hat{Y}}$ é a média dos valores preditos.

O viés ou erro sistemático médio (MBE), cuja expressão é dada por:

$$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i), \quad (2.6)$$

permite verificar se a predição é sistematicamente subestimada (valores negativos) ou superestimada (valores positivos), ou seja,

- i) $\hat{Y}_i - Y_i > 0 \implies \hat{Y}_i > Y_i$ (superestimação)
- ii) $\hat{Y}_i - Y_i < 0 \implies \hat{Y}_i < Y_i$ (subestimação)
- iii) $\hat{Y}_i - Y_i = 0 \implies \hat{Y}_i = Y_i$ (predição perfeita)

Essa medida representa a acurácia do modelo, conforme ALVES e VECCHIA (2011), o ideal é que esse valor tenda a zero, mas deve ser interpretado com cautela, uma vez que é consistentemente relacionado com a magnitude do erro.

A raiz do erro quadrático médio (RMSD), proposto por FOX (1981), consiste em avaliar a magnitude média dos erros de previsão, tem valor sempre positivo e quanto mais próximo de zero, maior a qualidade de predição em relação aos dados observados. Essa medida é capaz de apontar o quanto as predições se adequam bem aos dados observados, este índice não faz distinção quanto aos dados serem subestimados ou superestimados (JACOVIDES e KONTOYIANNIS, 1995).

A RMSD é dada por:

$$\text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}. \quad (2.7)$$

A desvantagem é que bastam alguns valores discrepantes para que ocorra aumento significativo em sua magnitude (STONE, 1993).

O índice de concordância (IA), de acordo com WILLMOTT (1982), é uma medida que reflete a precisão dos valores preditos em relação aos valores observados, cuja expressão é dada por:

$$\text{IA} = 1 - \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^n (|\hat{Y}_i - \bar{Y}| + |Y_i - \bar{Y}|)^2}. \quad (2.8)$$

Esse índice pode assumir valores entre 0 e 1 e quanto mais próximo de um, melhor é a concordância entre valores observados e preditos. BLAIN e BRUNINI (2005) citam que o índice de concordância quantifica numericamente a exatidão entre os valores estimados e observados de uma determinada variável. A última medida estatística é denominada eficiência da modelagem (EM) e averigua quanto as estimativas ou modelos estão correlacionados e eficientes entre si (DUNCAN, 1975). Foi proposta por, NASH e SUTCLIFFE (1970) dada por:

$$\text{EM} = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}. \quad (2.9)$$

Para um ajuste perfeito entre os dados preditos e observados o valor dessa estatística deve ser igual a 1, quando o resultado for menor ou igual a zero significa que o uso da média dos dados observados é melhor que os valores preditos. Conforme SILVA (2006), quando o valor de EM resultar em um valor maior do que 0,75, o desempenho do modelo é considerado bom. Para valores de EM entre 0,36 e 0,75, o desempenho é considerado aceitável, enquanto que valores de EM inferiores a 0,36 fazem com que o modelo seja julgado como inaceitável.

KOBAYASHI e SALAM (2000) apresentam uma medida que é $QMEP = (RMSE)^2$, ou seja, trabalham com o quadrado médio do erro de predição (QMEP) dividindo esta medida em diferentes componentes, conforme as equações:

$$QMEP = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 = SB + SDDS + LCS.$$

$$QMEP = (\bar{\hat{Y}} - \bar{Y})^2 + \frac{1}{n} \sum_{i=1}^n [(\hat{Y}_i - \bar{\hat{Y}}) - (Y_i - \bar{Y})]^2. \quad (2.10)$$

O primeiro termo do lado direito da expressão (2.12), que será denotado como SB, representa o viés entre a média dos valores preditos e a média dos valores observados da variável resposta, ou seja,

$$SB = (\bar{\hat{Y}} - \bar{Y})^2.$$

O segundo termo do lado direito da expressão (2.12), que será denotada como MSV, é a diferença entre os desvios das predições e observações em relação as suas respectivas médias, dado por:

$$MSV = \frac{1}{n} \sum_{i=1}^n [(\hat{Y}_i - \bar{\hat{Y}}) - (Y_i - \bar{Y})]^2. \quad (2.11)$$

Valor alto para esta componente (MSV) indica que o modelo é falho em representar a variabilidade das observações em torno da média.

A componente MSV, por sua vez, pode ser reescrita como a soma de duas outras componentes, dadas por:

$$MSV = (SD_s - SD_m)^2 + 2SD_s SD_m (1 - r), \quad (2.12)$$

em que: SD_s é o desvio padrão dos valores preditos; SD_m é o desvio padrão dos valores observados e r é o coeficiente de correlação entre os valores preditos e observados.

As manipulações algébricas para dedução da equação (2.12) a partir da equação (2.11) são apresentadas no Apêndice A detalhadamente.

As expressões para os desvios padrões, dos valores preditos (SD_s) e observados (SD_m), e do coeficiente de correlação de Pearson (r) entre esses valores, são dadas por:

$$SD_s = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2},$$

$$SD_m = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2},$$

$$r = \frac{\left[\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})(Y_i - \bar{Y}) \right]}{SD_m SD_s}.$$

O primeiro termo da equação (2.12), denotado por SDSD, fornece o grau de oscilação entre a predição e os dados observados, ou seja,

$$SDSD = (SD_s - SD_m)^2. \quad (2.13)$$

Um alto valor dessa componente indica que o modelo falha para simular o grau de oscilação entre as n medições.

O segundo termo da equação (2.12), denotada por LCS, representa a falta de correlação positiva ponderada pelos desvios padrões, denotada por:

$$LCS = 2SD_s SD_m (1 - r). \quad (2.14)$$

Um valor alto para essa componente indica que o modelo falha para simular um padrão da ao longo das n medições.

Com todos os termos acima combinados, o QMEP pode ser reescrito como:

$$QMEP = SB + SDSD + LCS. \quad (2.15)$$

A equação (2.14) indica o papel do coeficiente de correlação, r , em LCS e consequentemente em QMEP. Com todos os outros componentes fixos, um r maior reduziria QMEP e, consequentemente, aumentará a precisão do modelo. No entanto, r é apenas um dos componentes da QMEP, há casos em que as outras componentes podem ser mais influentes. Por exemplo, quando o SB é o componente principal do QMEP, maximizando r não aumenta a precisão do modelo. Na componente MSV nota-se que SDSD e LCS não são totalmente independentes, pois compartilham os mesmos componentes, SD_s e SD_M . Assim, se o desvio padrão dos valores preditos for maior que o desvio padrão dos valores observados irá aumentar o SDSD e LCS.

OLIVEIRA (2016) desenvolveu um estudo de concordância entre dois modelos para saber se a evapotranspiração obtida com o uso do irrigâmetro concorda com a evapotranspiração obtida pelo método padrão de Penman Monteith FAO 56, usando o coeficiente de correlação linear; o coeficiente de determinação; o índice de concordância e o índice de desempenho em dados de evapotranspiração. O estudo de simulação evidenciou que nenhuma das técnicas foi eficiente para indicar, sozinhas, a concordância dos dados simulados e que o índice de concordância e o índice de desempenho se mostraram mais eficientes no estudo. Ainda segundo a autora é possível notar que poucos são os trabalhos

que fazem uso do QMEP ou de sua raiz quadrada no complemento do estudo de índices de concordância IA SAIDI *et al.* (2015), MENGUE e FONTANA (2015), ALENCAR *et al.* (2011), PICOLI *et al.* (2009) e ANDRADE JUNIOR *et al.* (2003). Destes trabalhos apenas dois (PICOLI *et al.*, 2009) e (ANDRADE JUNIOR *et al.*, 2003) fizeram o desdobramento do erro em seus componentes sistemáticos e não-sistemáticos.

3 MATERIAIS E MÉTODOS

3.1 Banco de dados

Os dados analisados para o ajuste dos modelos foram extraídos de BRITO e SILVA (1983), um experimento em que foram mensurados o comprimento (cm), o diâmetro (cm) e o volume (cm^3) dos frutos de cacau. As medidas de comprimento e diâmetro foram obtidas por meio de um paquímetro, enquanto que o volume foi determinado pelo deslocamento do nível de água da coluna líquida. Nesse experimento temos a idade exata do fruto em alguns experimentos essa idade é aproximada, já que nesse caso foi utilizado o processo de polinização artificial.

Os frutos do cacau, 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 alfisol¹ da série germoplasma.

Na Tabela 3.1 são apresentados idade, o número de frutos analisados, a média para as medidas de comprimento, diâmetro e volume, respectivamente.

Tabela 3.1. Idade, número de frutos avaliados, médias das medidas de comprimento, diâmetro e volume do fruto do Cacauzeiro.

Idade (dias)	N de Frutos	Comprimento (cm)	Diâmetro (cm)	Volume (cm^3) ³
30	50	1,92	0,82	0,86
45	40	3,04	1,32	2,82
60	30	4,50	1,94	7,91
75	30	5,51	2,51	17,77
90	20	8,06	4,03	69,70
105	20	10,01	6,00	191,50
120	20	10,16	7,12	288,00
135	20	10,61	8,26	395,75
150	15	10,61	8,56	421,00
165	15	10,64	8,49	410,00
180	16	10,37	8,30	365,08

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, tomaram-se 40 frutos na segunda coleta, 30 frutos na terceira e na quarta coleta. 20 frutos, na quinta, sexta, sétima e oitava coleta, 15 frutos na nona e décima

¹São solos com maior conteúdo de argila, mas sem húmus, são de florestas alcalinas, algumas vezes ácidas e são usadas para o cultivo de cereais e pastagens.

coleta e por fim 16 frutos na décima primeira coleta. Os frutos coletados eram colocados em sacos plásticos e imediatamente levados ao laboratório, onde eram tomadas as medidas de crescimento.

3.2 Ajuste dos modelos

Os modelos não lineares: Logístico, Gompertz e Von Bertalanffy foram ajustados para as medidas de comprimento e diâmetro.

O modelo Logístico é adequado à representação de dados de crescimento de organismos vivos, onde o crescimento inicial é exponencial e, após o ponto de inflexão, assintótico. Sua expressão é dada por:

$$Y_i = \frac{\beta_0}{[1 + \exp(-\beta_1 - \beta_2 X_i)]} + \epsilon_i. \quad (3.1)$$

O modelo Gompertz é uma função monotonicamente crescente, cuja expressão é dada por:

$$Y_i = \beta_0 \exp(-\exp(\beta_1 - \beta_2 X_i)) + \epsilon_i. \quad (3.2)$$

O modelo de Von Bertalanffy é bastante empregado em estudos de crescimento de peixes e é definido por:

$$Y_i = \beta_0 (1 - \exp(-\beta_1 - \beta_2 X_i))^3 + \epsilon_i. \quad (3.3)$$

Nas expressões (3.1), (3.2) e (3.3), $i = 1, 2, \dots, n$, Y_i é a variável resposta, X_i uma medida de tempo, β_0 representa a assíntota horizontal, ponto onde se estabiliza o crescimento; β_1 não apresenta uma interpretação direta, estando relacionado ao intercepto da curva; β_2 é a taxa de velocidade do crescimento e ϵ_i é o erro experimental.

Na análise de regressão se faz necessário obter as estimativas dos parâmetros do modelo, em sua maioria é mais prático encontrar as estimativas de mínimos quadrados por procedimentos de procura numérica direta, tais como método de Newton e o método de Newton-Raphson, do que, inicialmente, obter as equações normais e, então, usar métodos numéricos para encontrar a solução dessas equações iterativamente. O método de Gauss-Newton, também conhecido como método da linearização, usa uma expansão em série de Taylor de primeira ordem para aproximar o modelo de regressão não linear com termos lineares e, então, aplica mínimos quadrados ordinário para estimar os parâmetros. Iterações desses passos geralmente conduzem a uma solução para o problema de regressão não linear. Para escolher qual modelo melhor se ajustou ao banco de dados utilizamos o critério de informação de Akaike (AIC), critério de informação Baysiano e o desvio padrão residual.

Com os valores estimados dessas características por meio dos modelos propostos, o volume do fruto foi obtido utilizando a fórmula do esferoide prolato, e comparado ao

volume encontrado no experimento determinado por meio do deslocamento do nível de água. O volume do esferoide prolato é dado por:

$$V = \frac{4}{3}\pi \left(\frac{a}{2}\right) \left(\frac{b}{2}\right)^2, \quad (3.4)$$

em que: a é o comprimento do fruto estimado por meio dos modelos Logístico, Gompertz e Von Bertalanffy e b é o diâmetro do fruto estimado utilizando os mesmos modelos.

Para avaliar a qualidade de ajuste de cada modelo na predição do volume do fruto foram utilizadas quatro medidas estatísticas descritas na Seção (2.7), sendo que uma delas, o quadrado médio do erro de predição ($QMEP = RMSD^2$), foi dividida em três componentes.

As análises foram realizadas com o auxílio do *software* estatístico R (R CORE TEAM, 2017). Para o ajuste dos modelos Logístico, Gompertz e Von Bertalanffy foi utilizado a função *nls* do pacote *stats*. O ajuste foi realizado para as médias dos dados, comprimento e diâmetro, de cada dia de coleta, o valor inicial para o parâmetro β_0 foi obtido utilizando o maior valor de cada medida mais um.

3.3 Técnicas de diagnóstico

Para verificar se as suposições do modelo estão satisfeitas e identificar características dos dados foram utilizadas técnicas de diagnóstico gráficas que permitem avaliar a adequação do modelo proposto. Nesse trabalho é apresentado o gráfico dos resíduos padronizados e o Q-Q plot com envelope simulado.

Uma das técnicas de diagnóstico para modelos de regressão é a análise de resíduos. Resíduo é a diferença entre o valor observado e o valor ajustado, resíduos com alto valor indicam que o modelo proposto não foi bem ajustado aos dados.

Assim, os resíduos padronizados são definidos por:

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}}, \quad i = 1, 2, \dots, n,$$

em que h_{ii} é o i -ésimo elemento da matriz de projeção ortogonal H .

As observações com resíduo padronizado fora do intervalo $-3 \leq r_i \leq 3$ são considerados *outliers*.

Nesse tipo de gráfico, espera-se que os pontos estejam dispersos aleatoriamente entre os limites do envelope, a presença de pontos externos ao envelope ou de pontos internos apresentando padrões sistemáticos podem indicar problemas de ajuste.

O teste de Shapiro Wilk investiga se uma amostra aleatória provém de uma distribuição normal.

4 RESULTADOS

Nesta seção serão apresentados os resultados do ajuste dos modelos Logístico, Gompertz e Von Bertalanffy aos dados de comprimento e diâmetro do fruto do cacau, bem como os resultados da predição do volume do fruto utilizando os valores estimados de comprimento e diâmetro por meio dos modelos discutidos na seção 3.

Nas Figuras 4.1 e 4.2, são apresentados os *box-plot* para as medidas de comprimento e diâmetro, respectivamente, em cada uma das coletas realizadas. Em ambas as figuras é possível perceber algumas características em relação à distribuição dos dados, ou seja, não foram detectados valores discrepantes nas amostras, as amplitudes dos dados ao longo das coletas são aparentemente diferentes, indicando que no início das coletas os frutos são mais homogêneos entre si no que se refere ao seu comprimento e diâmetro, com o passar das coletas vão ficando heterogêneas e que a posição centralizada da linha mediana nas caixas indica uma distribuição simétrica dos dados na maioria das coletas. Uma pequena distorção dessa simetria pode ser visualizada no box-plot para o comprimento dos frutos coletados 60 e 75 dias após a polinização (Figura 4.1) em que a linha mediana está deslocada em direção ao primeiro quartil da amostra.

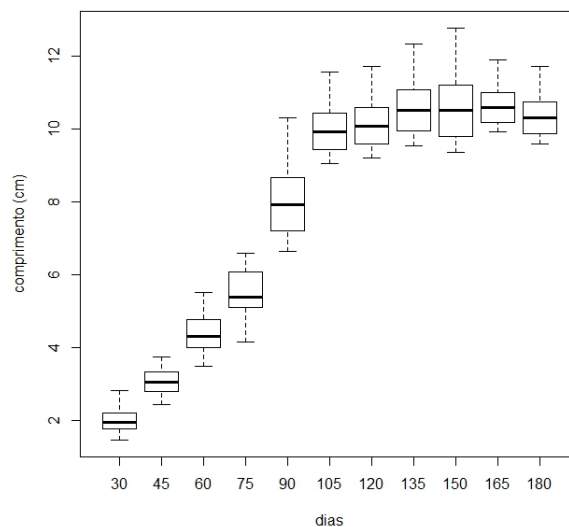


Figura 4.1. Box-plot dos dados do comprimento, em centímetros, de cacau ao longo dos dias após a polinização.

O ajuste dos modelos Logístico, Gompertz e Von Bertalanffy foram realizados para as médias dos valores de comprimento e diâmetro dos frutos do cacau e estão representados nas Figuras 4.3 e 4.4. Nestas figuras, é possível perceber que em média o comportamento dos dados aproxima-se de uma curva sigmoide e que, em relação às dimensões dos frutos, aparentemente, há uma estabilização em torno de 11cm de comprimento e 10cm de diâmetro. Também é possível verificar nesta figura que por volta de 60 dias após

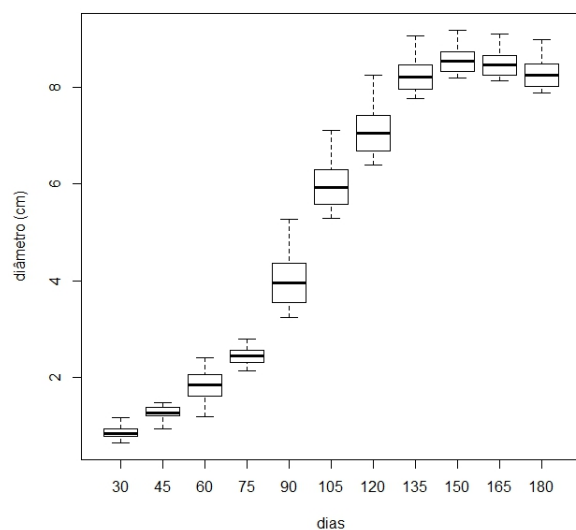


Figura 4.2. Box-plot dos dados diâmetro, em centímetro, de cacau ao longo dos dias após a polinização.

a polinização o crescimento do fruto, em comprimento, passa de um crescimento acelerado para um crescimento desacelerado o que coincide com o ponto de inflexão da curva. Em relação ao crescimento em diâmetro o ponto de inflexão aparentemente acontece por volta de 80 dias após a polinização.

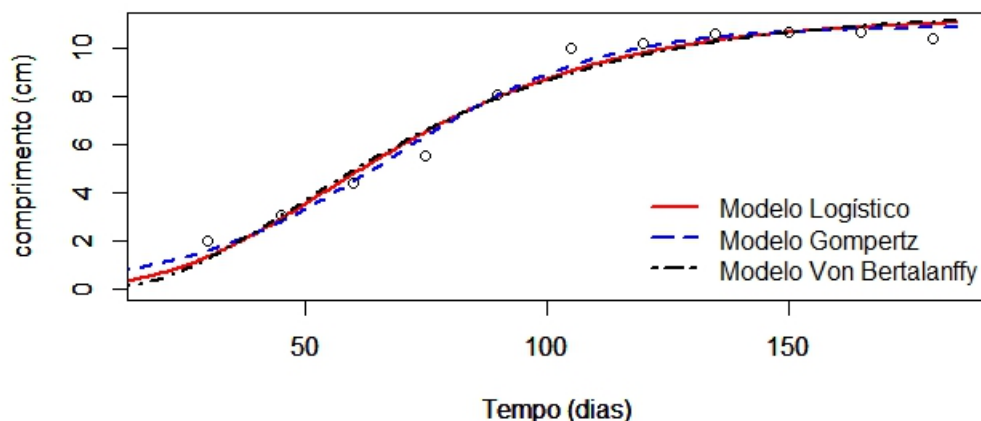


Figura 4.3. Médias e ajuste do modelo Logístico, Gompertz e Von Bertalanffy dos dados do comprimento, em centímetro, de cacau em função dos dias.

Nas Tabelas 4.1, 4.2 e 4.3, são apresentados os valores das estimativas dos parâmetros bem como os avaliadores da qualidade de ajuste dos modelos Logístico, Gompertz e Von Bertalanffy, respectivamente. Para testar a normalidade dos resíduos, foi utilizado

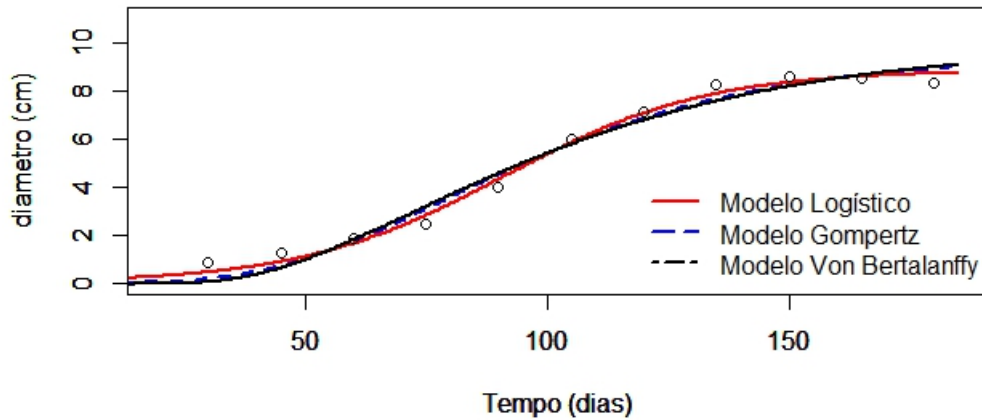


Figura 4.4. Médias e ajuste dos modelos Logístico, Gompertz e Von Bertalanffy dos dados de diâmetro, em centímetro, de cacau em função dos dias.

o teste de Shapiro Wilk, como o p -valor do teste foi maior que o nível de 5% de significância, não rejeitamos a hipótese de normalidade dos resíduos, ou seja, conclui-se que os erros são normalmente distribuídos. É interessante notar que os três modelos de crescimento apresentam baixos valores para o desvio padrão residual (RSD), valores baixos para o critério de informação de Akaike (AIC) e informação Bayesiano (BIC), sendo o modelo Logístico o de menor valor.

Tabela 4.1. Estimativas dos parâmetros (erro padrão) e avaliadores da qualidade de ajuste do modelo Logístico, na descrição de comprimento e diâmetro dos frutos do cacau-eiro.

Parâmetros	Comprimento	Diâmetro
β_0	10,89 (0,2823)	8,84 (0,2655)
β_1	-3,15 (0,3244)	-4,318 (0,3805)
β_2	0,05 (0,0051)	0,05 (0,0047)
Avaliadores	Comprimento	Diâmetro
RSD	0,469	0,3473
AIC	19,06	12,45
BIC	20,65	14,04
Shapiro-Wilk	p=0,8209	p=0,1372

Para variável comprimento do fruto, verifica-se em relação ao parâmetro β_0 , ou seja, o comprimento assintótico do fruto, que os três modelos ajustados forneceram estimativas bastante próximas entre si, variando de 10,89 cm para o modelo Logístico (Tabela 4.1) a 11,50 cm para o modelo Von Bertalanffy (Tabela 4.3), o que do ponto de vista prático representa uma diferença muito pequena, em torno de 0,6cm na estimativa do comprimento assintótico do fruto. Porém, a estimativa do parâmetro β_0 obtida pelo ajuste do

Tabela 4.2. Estimativas dos parâmetros (erro padrão) e avaliadores da qualidade de ajuste do modelo Gompertz, na descrição de comprimento e diâmetro dos frutos do cacauero.

Parâmetros	Comprimento	Diâmetro
β_0	11,26 (0,5093)	9,47 (0,6694)
β_1	1,66 (0,2662)	2,19 (0,3549)
β_2	0,03 (0,0049)	0,03 (0,0051)
Avaliadores	Comprimento	Diâmetro
RSD	0,6351	0,5436
AIC	25,72	22,30
BIC	27,32	23,89
Shapiro-Wilk	p=0,9978	p=0,1817

Tabela 4.3. Estimativas dos parâmetros (erro padrão) e avaliadores da qualidade de ajuste do modelo Von Bertalanffy, na descrição de comprimento e diâmetro dos frutos do cacauero.

Parâmetros	Comprimento	Diâmetro
β_0	11,50 (0,6697)	9,95 (1,0371)
β_1	-0,09 (0,2379)	-0,44 (0,3369)
β_2	0,025 (0,0048)	0,021 (0,0052)
Avaliadores	Comprimento	Diâmetro
RSD	0,7058	0,6303
AIC	28,05	25,26
BIC	29,64	27,15
Shapiro-Wilk	p=0,9948	p=0,3074

modelo Logístico é mais precisa quando comparada às estimativas obtidas pelo ajuste dos demais modelos, o que pode ser verificado por meio do erro padrão da estimativa do parâmetro que foi menor para o modelo Logístico. Ainda em relação ao comprimento do fruto é possível verificar que embora o modelo Logístico, em comparação aos demais modelos, tenha apresentado a menor estimativa para o comprimento assintótico do fruto, o mesmo resultou na maior estimativa do parâmetro β_2 que representa a taxa de crescimento do fruto.

Já para a variável diâmetro, o parâmetro β_0 varia de 8,84 (modelo Logístico) até 9,95 (modelo Von Bertalanffy), sendo o erro padrão do modelo Logístico menor do que os demais. Como os parâmetros β_0 e β_2 , possuem correlação negativa, ou seja, um aumenta e outro diminui e um representa o crescimento assintótico e o outro a taxa de crescimento do fruto, é possível verificar, vide (Tabelas 4.1, 4.2 e 4.3), que a taxa de crescimento do modelo Logístico é maior comparada aos demais modelos, mesmo possuindo a menor estimativa para o diâmetro assintótico.

A variedade de cacau estudada aqui não foi comercializada, no entanto ALEXANDRE *et al.* (2015) estudou a caracterização de frutos de clones de cacaueros na região litorânea de São Mateus e as medidas de comprimento e diâmetro encontradas variam

entre 13 à 23 cm e 9 à 11 cm respectivamente. Valores mais aproximados com os apresentados nas Tabelas 4.1, 4.2 e 4.3, foram encontrados por SILVA *et al.* (2013) onde o comprimento variava entre 9,16 à 14,67 cm e o diâmetro dos frutos variou entre 5,32 à 7,44 cm .

Na Tabela 4.4 são apresentadas as coordenadas das abscissas dos pontos de inflexão das curvas com os respectivos intervalos de 95% de confiança, para as medidas de comprimento e diâmetro, obtidas por meio de cada um dos modelos estudados. Para a medida de comprimento no modelo Logístico, Gompertz e Von Bertalanffy o fruto atinge a metade do seu desenvolvimento quando a taxa é máxima, que foi em 68 dias, 55 dias e 48 dias, respectivamente, depois a taxa de crescimento reduz até que estabilize e as medidas atinjam o valor assintótico. Já para a medida de diâmetro essa taxa ocorre em 91 dias, 79 dias e 62 dias para os modelos Logístico, Gompertz e Von Bertalanffy. O modelo Logístico apresentou o menor intervalo de confiança, sendo o modelo que descreve melhor o comportamento das medidas de comprimento e diâmetro do cacau.

Tabela 4.4. Coordenada da abscissa do ponto de inflexão estimadas com os modelos de crescimento Logístico, Gompertz e Von Bertalanffy.

Medidas	Intervalo do ponto de inflexão
Modelo Logístico	
Comprimento	[62,5 - 73,1]
Diâmetro	[83,5- 98,0]
Modelo Gompertz	
Comprimento	[49,8 - 59,9]
Diâmetro	[71,3 - 86,9]
Modelo Von Bertalanffy	
Comprimento	[35,33 - 60,5]
Diâmetro	[44,9-79,4]

Na Tabela 4.5 temos as medidas estatísticas da qualidade de predição do volume do fruto do cacauero. Com relação à variabilidade relativa dos avaliadores, temos o MBE, conforme ALVES e VECCHIA (2011) quanto mais próximo de zero, melhor, sendo negativo para os três modelos mostrando que o volume pelo esferoide prolato está subestimando o valor verdadeiro.

Já o índice de concordância (IA) pode assumir valores entre 0 e 1, quanto mais próximo de 1, maior é a concordância entre os valores observados e preditos, nos três modelos temos valores próximos a um, sendo o modelo Logístico o adequado.

A eficiência da modelagem (EM) assume valores próximos a um e valores próximos entres os modelos, o modelo Logístico se aqueda melhor com relação a eficiência da modelagem, segundo SILVA (2006) para um valor maior do que 0,75 o desempenho do modelo é considerado bom.

O QMEP foi menor para o modelo Logístico, medida que aponta quanto as pre-

Tabela 4.5. Medidas estatísticas da qualidade de predição do volume do fruto do cacau-eiro.

Avaliadores	Modelo Logístico	Modelo Gompertz	Von Bertalanffy
MBE	-4,7744	-5,7696	-6,61
IA	0,9922	0,9841	0,9797
EM	0,9696	0,9382	0,9217
RMSD	30,27548	43,17877	48,61031
QMEP	916,6045	1864,406	2362,962
SB	22,79507	33,28868	43,7587
SDSD	8,042043	5,30088	5,069939
LCS	885,7674	1825,816	2314,134

dições se adequam aos dados observados, essa medida apresentada por KOBAYASHI e SALAM (2000) é subdividida em outras três.

A medida SB indica que houve superestimação do volume do fruto do cacau usando o esferoide prolato, o que acaba refletindo também na medida SDSD, visto que essa fornece a oscilação dos desvios padrões dos valores preditos e observados, já a medida LCS indica que o modelo precisa de ajustes no padrão de previsão com relação ao conjunto de dados, o que pode ser explicado pelo fato do fruto perder volume no final do processo de maturação, nesse caso o modelo Logístico apresentou todas essas medidas em porcentagem menor.

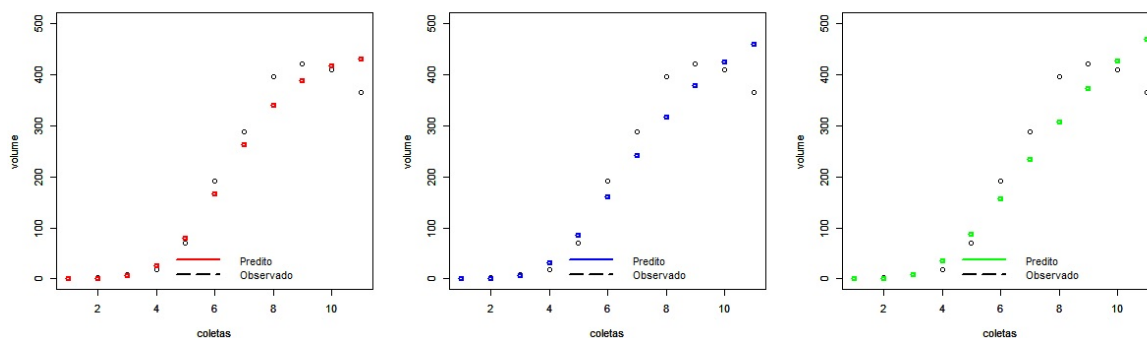


Figura 4.5. Volume observado e estimado do cacau em cm^3 para os modelos (a) Logístico, (b) Gompertz e (c) Von Bertalanffy em função do tempo após a polinização

Na Figura 4.5 são apresentados o diagrama de dispersão do volume estimado pelo esferoide prolato e o volume obtido experimentalmente, para cada um dos modelos analisados. Segundo FREITAS (2005) os modelos de crescimento resumem em três ou quatro parâmetros características da população, pois alguns dos parâmetros possuem interpretação biológica. Conforme exposto por NASCIMENTO (2014) pelo fato de que o fruto perde líquido no fim do processo de maturação e conseqüentemente há a perda de volume. Os modelos propostos não conseguem captar esse decaimento no fim do processo. A partir da análise de regressão não linear entre os volumes dos frutos medidos no experimento e

estimado pela fórmula matemática, o modelo Logístico apresentou melhor resultado, já que o mais importante é obter informações sobre todo o ciclo de crescimento. NETO e REINHARDT (2003) utilizaram um método não destrutivo para medir o volume do fruto da mangueira, utilizando as medidas de comprimento, circunferência e volume. Comparando esse valor com o volume obtido com a água deslocada, os resultados sugerem a possibilidade de estimar o volume do fruto a partir das medidas biométricas, evitando assim a retirada do fruto.

As Figuras 4.6 e 4.7 nos fornecem os gráficos dos resíduos padronizados para as medidas de comprimento e diâmetro. Mesmo apresentando uma tendência é importante resaltar que com relação aos pressupostos, os resíduos não são autocorrelacionados e não ocorre heterogeneidade de variâncias já que as medidas de crescimento não foram tomadas na mesma unidade experimental. A Tabela 4.6 apresenta as medidas de não linearidade intrínseca, o modelo Logístico foi o que apresentou menor valor, sendo ele o que mais se aproxima da linearidade.

As Figuras 4.8 e 4.9 avaliam a qualidade do ajuste e se fez necessário uma análise de diagnóstico, para testar a normalidade dos resíduos do modelo, observa-se no gráfico normal de probabilidade que todos os pontos encontram-se dentro do envelope de simulado, sendo comprovada por meio do teste de Shapiro-Wilk, não rejeitando a hipótese de que há normalidade residual considerando o nível de 5% de significância.

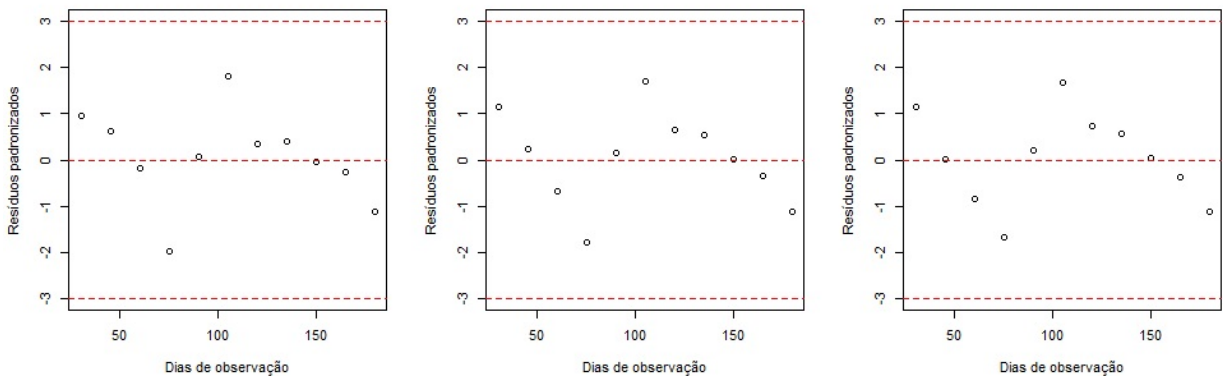


Figura 4.6. Resíduos padronizados para a medida de comprimento do cacau para os modelos Logístico, Gompertz e Von Bertalanffy, respectivamente

Tabela 4.6. Medida de não linearidade para as medidas de comprimento e diâmetro

	Comprimento	Diâmetro
Modelo	Intrínseco	Intrínseco
Logístico	0,1562	0,1510
Gompertz	0,1771	0,2923
Von Bertalanffy	0,1762	0,5563

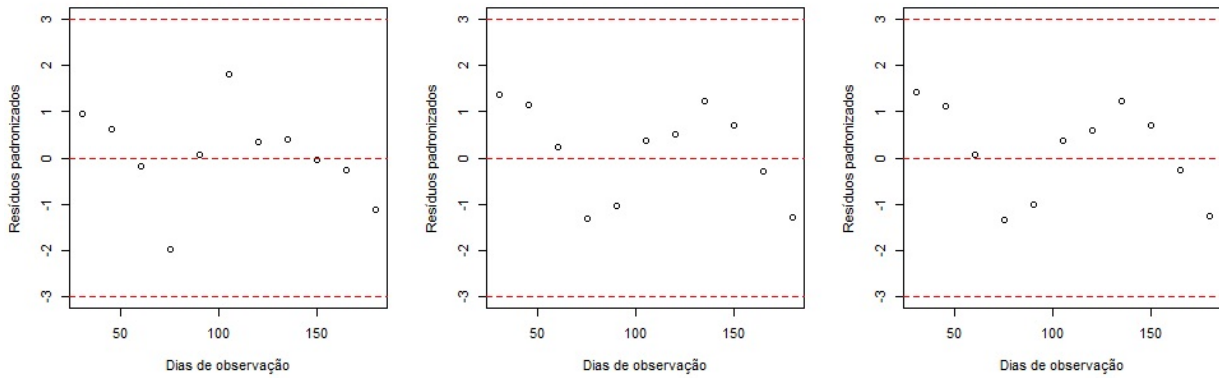


Figura 4.7. Resíduos padronizados para a medida de diâmetro do cacau para os modelos Logístico, Gompertz e Von Bertalanffy, respectivamente

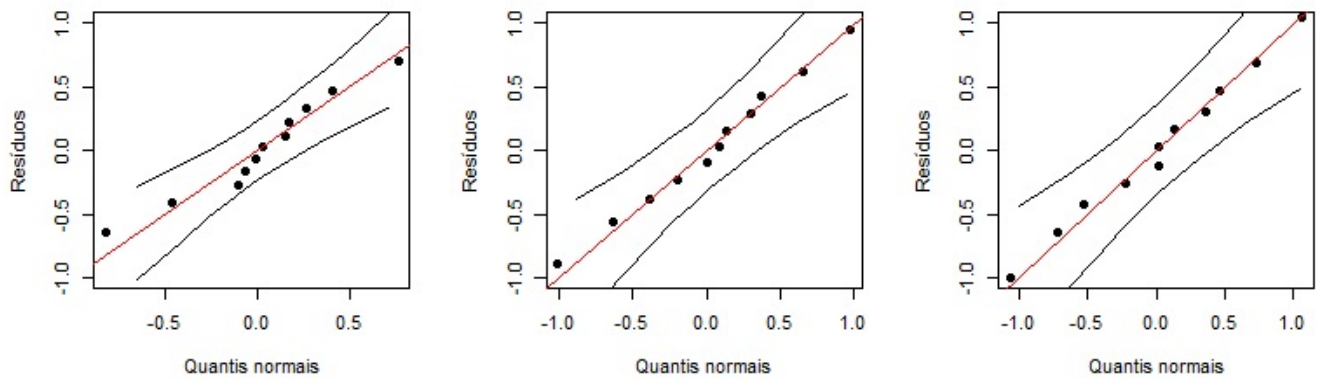


Figura 4.8. Gráfico Normal de probabilidade para a medida de comprimento do cacau para os modelos Logístico, Gompertz e Von Bertalanffy, respectivamente.

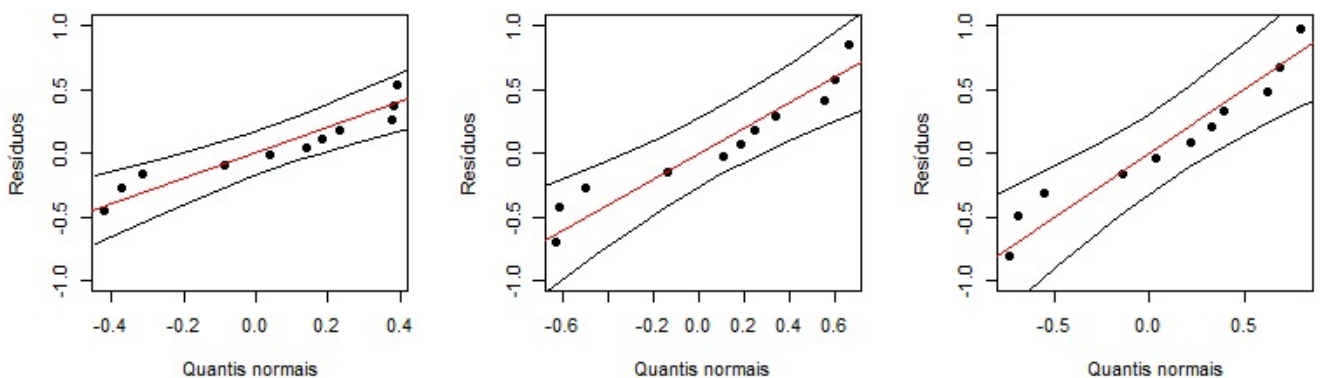


Figura 4.9. Gráfico Normal de probabilidade para a medida de diâmetro do cacau para os modelos Logístico, Gompertz e Von Bertalanffy, respectivamente.

5 CONCLUSÃO

As medidas estatísticas utilizadas no presente trabalho mostraram-se eficazes para avaliar a qualidade do ajuste e predição do volume do cacau. Segundo os avaliadores utilizados o modelo Logístico foi o que apresentou os melhores ajustes aos dados de comprimento e diâmetro do fruto de cacau bem como a melhor capacidade de predição do volume desses frutos.

Em relação as estimativas dos parâmetros é possível observar que, para o parâmetro β_0 foi obtida uma estimativa de aproximadamente 10,89 cm para comprimento e 8,84 cm para diâmetro representando a medida assintótica do fruto. Uma medida importante é o ponto de inflexão das medidas, indicando metade do crescimento, sendo 5,45 cm para comprimento e 4,42 cm para diâmetro. Com relação a interpretabilidade dos parâmetros em 63 dias ocorreu a velocidade máxima do crescimento por dia da medida de comprimento e 87 dias para a medida do diâmetro e que após esses dias houve uma desaceleração no crescimento das medidas do fruto.

Considerando a bondade do ajuste do modelo Logístico aos dados de comprimento e diâmetro do fruto do cacau foi possível, por meio dos valores estimados dessas características, predizer, de forma satisfatória, o volume dos frutos sem a necessidade de retirada do fruto antes do processo de maturação se completar, evitando assim desperdício econômicos e redução na produção da cultura.

REFERÊNCIAS

- AKAIKE, H., 1974 A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* **19**: 716–723.
- ALENCAR, L. D., G. C. SEDIYAMA, H. S. WANDERLEY, T. S. ALMEIDA, e R. C. DELGADO, 2011 Avaliação de métodos de estimativa da evapotranspiração de referência para três localidades no norte de minas gerais. *Revista Engenharia na Agricultura* **19**: 437–449.
- ALEXANDRE, R. S., K. CHAGAS, H. I. MARQUES, P. R. COSTA, *et al.*, 2015 Caracterização de frutos de clones de cacauzeiros na região litorânea de são mateus, es. *Revista Brasileira de Engenharia Agrícola e Ambiental-Agriambi* **19**.
- ALVES, E. D. L. e F. A. S. VECCHIA, 2011 Análise de diferentes métodos de interpolação para a precipitação pluvial no estado de goiás-doi: 10.4025/actascihumansoc. v33i2. 13815. *Acta Scientiarum. Human and Social Sciences* **33**: 193–197.
- ANDRADE JUNIOR, A. D., E. A. BASTOS, P. C. SENTELHAS, e A. D. SILVA, 2003 Métodos de estimativa da evapotranspiração de referência diária para parnaíba e teresina, piauí. *Revista Brasileira de Agrometeorologia, Santa Maria* **11**: 63–68.
- BATES, D. M. e D. G. WATTS, 1980 Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 1–25.
- BATISTA, A. P. S. A., 2009 Chocolate: sua história e principais características .
- BATISTA, N. L. e L. M. D. VIERO, 2014 Cacau, um dos grandes ciclos econômico do brasil .
- BEALE, E., 1960 Confidence regions in non-linear estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 41–88.
- BLAIN, G. C. e O. BRUNINI, 2005 Avaliação e adaptação do índice de severidade de seca de palmer (pdsi) e do índice padronizado de precipitação (spi) às condições climáticas do estado de são paulo. *Bragantia* **64**.
- BRITO, C. C. R., J. A. A. DA SILVA, R. L. C. FERREIRA, E. DE SOUZA SANTOS, e I. FERRAZ, 2007 Modelos de crescimento resultantes da combinação e variações dos modelos de chapman-richards e silva-bailey aplicados em *leucaena leucocephala* (lam.) de wit. *Ciência Florestal* **17**: 175–185.
- BRITO, I. e C. SILVA, 1983 Medidas biométricas do fruto do cacauzeiro durante seu desenvolvimento. *Sitientibus, Feira de Santana* **2**: 59–66.

- CUENCA, M. A. G. e C. C. NAZÁRIO, 2004 Importância econômica e evolução da cultura do cacau no brasil e na região dos tabuleiros costeiros da bahia entre 1990 e 2002. Embrapa Tabuleiros Costeiros. Documentos .
- DEMPSTER, A. P., N. M. LAIRD, e D. B. RUBIN, 1977 Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)* pp. 1–38.
- DETMANN, E., S. VALADARES FILHO, L. T. HENRIQUES, D. PINA, M. F. PAULINO, A. L. R. MAGALHÃES, D. FIGUEIREDO, M. PORTO, e M. L. CHIZZOTTI, 2007 Reparametrização do modelo baseado na lei de superfície para predição da fração digestível da fibra em detergente neutro em condições brasileiras. *Revista Brasileira de Zootecnia* **36**: 155–164.
- DUNCAN, O. D., 1975 *Introduction to structural equation models*. academic press, London.
- EFRAIM, P., 2009 Contribuição à melhoria de qualidade de produtos de cacau no brasil, por meio da caracterização de derivados de cultivares resistentes à vassoura-de-bruxa e de sementes danificadas pelo fungo .
- FIGUEIREDO FILHO, D., F. NUNES, E. C. DA ROCHA, M. L. SANTOS, M. BATISTA, e J. A. SILVA JÚNIOR, 2011 O que fazer e o que não fazer com a regressão: pressupostos e aplicações do modelo linear de mínimos quadrados ordinários (mqo). *Revista Política Hoje* **20**.
- FITZHUGH, H., 1976 Analysis of growth curves and strategies for altering their shape. *Journal of animal Science* **42**: 1036–1051.
- FOX, D. G., 1981 Judging air quality model performance. *Bulletin of the American Meteorological Society* **62**: 599–609.
- FREITAS, A. R., 2005 Curvas de crescimento na produção animal. *R. Bras. Zootec* **34**: 786–795.
- HARTLEY, H. O., 1961a The modified gauss-newton method for the fitting of non-linear regression functions by least squares. *Technometrics* **3**: 269–280.
- HARTLEY, H. O., 1961b The modified gauss-newton method for the fitting of non-linear regression functions by least squares. *Technometrics* **3**: 269–280.
- HERMÉ, P., 2007 *El chocolate*. Larousse.
- JACOVIDES, C. e H. KONTOYIANNIS, 1995 Statistical procedures for the evaluation of evapotranspiration computing models. *Agricultural Water Management* **27**: 365–371.

- KOBAYASHI, K. e M. U. SALAM, 2000 Comparing simulated and measured values using mean squared deviation and its components. *Agronomy Journal* **92**: 345–352.
- LAIRD, A. K. *et al.*, 1965 Dynamics of relative growth. *Growth* **29**: 249–263.
- MARTINS, J. M. *et al.*, 2011 Melhoria da qualidade de cacau. Ilhéus, BA: CEP-LAC/CENEX .
- MARTINS, M. C., 2013 *Análise de resíduos projetados em modelos não-lineares de regressão: uma aplicação em crescimento de frutos*. Ph.D. thesis, Universidade de São Paulo.
- MAZUCHELI, J. e J. A. ACHCAR, 2008 Algumas considerações em regressão não linear. *Acta Scientiarum. Technology* **24**: 1761–1770.
- MENGUE, V. P. e D. C. FONTANA, 2015 Avaliação da dinâmica espectro-temporal visando o mapeamento dos principais cultivos de verão no rio grande do sul. *Bragantia* **74**: 331–340.
- MISCHAN, M. M. e S. Z. DE PINHO, 2014 Modelos não lineares .
- NASCIMENTO, M. D. S., 2014 Modelos não lineares na descrição do crescimento de frutos de cacaueiros .
- NASH, J. E. e J. V. SUTCLIFFE, 1970 River flow forecasting through conceptual models part i: A discussion of principles. *Journal of hydrology* **10**: 282–290.
- NELDER, J., 1961 The fitting of a generalization of the logistic curve. *Biometrics* **17**: 89–110.
- NETO, M. C. T. e D. H. REINHARDT, 2003 Relações entre parâmetros de crescimento do fruto da manga cv. haden. *Rev. Bras. Frutic.* **25**.
- NOGUEIRA, B. L., 2015 *Processamento do cacau: avaliação do teor nutricional do chocolate e dos outros derivados do cacau*. Graduação, Escola de Engenharia de Lorena, Departamento de Engenharia Bioquímica.
- OLIVEIRA, E. A., 2016 *Métodos para análise de concordância: estudo de simulação e aplicação a dados de evapotranspiração*. Ph.D. thesis, Escola Superior de Agricultura Luiz de Queiroz.
- PICOLI, M. C. A., B. F. T. RUDORFF, R. RIZZI, A. GIAROLLA, *et al.*, 2009 Índice de vegetação do sensor modis na estimativa da produtividade agrícola da cana-de-açúcar. *Bragantia* **68**: 789–795.

- PINHO, S. Z. D., L. R. D. CARVALHO, M. M. MISCHAN, e J. R. D. S. PASSOS, 2014 Critical points on growth curves in autoregressive and mixed models. *Scientia Agricola* **71**: 30–37.
- PIRES, J. L. *et al.*, 2003 Avaliação quantitativa e molecular de germoplasma para o melhoramento do cacauero com ênfase na produtividade, qualidade de frutos e resistência a doenças .
- PRADO, T. K. L., T. V. SAVIAN, e J. A. MUNIZ, 2013 Ajuste dos modelos gompertz e logístico aos dados de crescimento de frutos de coqueiro anão verde. *Ciência Rural* **43**: 803–809.
- R CORE TEAM, 2017 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- SAIDI, D., Z. HAMEL, e A. ABABOU, 2015 Using pedotransfer functions to assess aggregate stability: Application to the lower chelif soils, algeria. *International Journal of Plant & Soil Science* .
- SCHWARZ, G. *et al.*, 1978 Estimating the dimension of a model. *The annals of statistics* **6**: 461–464.
- SILVA, J. S. F., 2006 Modelagem de equações estruturais: Apresentação de uma metodologia .
- SILVA, S. S. D., C. D. S. BARBOSA, A. P. L. D. MESSIAS, R. S. D. LIMA, A. S. D. SILVA, e R. F. CAMPOS, 2013 Biometria dos frutos e sementes do cacau (*theobroma cacao* sterculiaceae) nativo do acre. 65 Reunião Anual da Sociedade Brasileira para o Progresso da Ciência .
- SMYTH, G. K., 2006 in *Nonlinear Regression*, John Wiley & Sons, Ltd.
- STONE, R., 1993 Improved statistical procedure for the evaluation of solar radiation estimation models. *Solar energy* **51**: 289–291.
- TERRA, M., J. MUNIZ, e T. SAVIAN, 2010 Ajuste dos modelos logístico e gompertz aos dados de crescimento de frutos da tamareira-anã (*phoenix roebelenii* o'brien). *Magistra, Cruz das Almas* **22**: 1–7.
- VON BERTALANFFY, L., 1957 Quantitative laws in metabolism and growth. *Quarterly review of Biology* pp. 217–231.
- WILLMOTT, C. J., 1982 Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society* **63**: 1309–1313.

APÊNDICES

Apêndice A

Cálculos especificados de forma detalhada

As manipulações algébricas para dedução da equação (2.12) a partir da equação (2.11) são apresentadas com detalhes a seguir.

$$\begin{aligned}
 MSV &= \frac{1}{n} \sum_{i=1}^n [(\hat{Y}_i - \bar{\hat{Y}}) - (Y_i - \bar{Y})]^2 \\
 &= \frac{1}{n} \sum_{i=1}^n [(\hat{Y}_i - \bar{\hat{Y}})^2 - 2(\hat{Y}_i - \bar{\hat{Y}})(Y_i - \bar{Y}) + (Y_i - \bar{Y})^2] \\
 &= \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2 + \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 + \frac{1}{n} \sum_{i=1}^n [-2(\hat{Y}_i - \bar{\hat{Y}})(Y_i - \bar{Y})]
 \end{aligned}$$

Acrescentando e subtraindo na equação anterior

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

temos que:

$$\begin{aligned}
\text{MSV} &= \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 - 2 \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} + 2 \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \\
&\quad + \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 + \frac{1}{n} \sum_{i=1}^n [-2(\hat{Y}_i - \bar{Y})(Y_i - \bar{Y})] \\
&= \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} - \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \right)^2 + 2 \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \\
&\quad - \frac{2}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \bar{Y}) \\
&= \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} - \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \right)^2 + 2 \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \\
&\quad \left(1 - \left[\frac{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2}} \right] \right) \\
&= \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} - \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \right)^2 + 2 \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \\
&\quad \left(1 - \left[\frac{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \bar{Y})}{SD_s SD_m} \right] \right) \\
&= \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} - \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \right)^2 + 2 \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} (1 - r) \\
\text{MSV} &= (SD_s SD_m)^2 + 2SD_s SD_m (1 - r).
\end{aligned}$$