

GUILHERME ALVES PUIATTI

**TÉCNICA DE AGRUPAMENTO NA SELEÇÃO DE MODELOS DE REGRESSÃO
NÃO LINEARES PARA DESCRIÇÃO DO ACÚMULO DE MATÉRIA SECA EM
PLANTAS DE ALHO**

Dissertação apresentada à
Universidade Federal de Viçosa, como
parte das exigências do Programa de
Pós-Graduação em Estatística Aplicada
e Biometria, para obtenção do título de
Magister Scientiae.

VIÇOSA
MINAS GERAIS – BRASIL
2014

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

P979t Puiatti, Guilherme Alves, 1988-
2014 Técnica de agrupamento na seleção de modelos de
regressão não lineares para descrição do acúmulo de matéria
seca em plantas de alho / Guilherme Alves Puiatti. – Viçosa,
MG, 2014.
viii, 61 f. : il. ; 29 cm.

Inclui apêndices.

Orientador: Paulo Roberto Cecon.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 45-51.

1. Análise de agrupamento. 2. Alho. 3. Análise de
regressão. 4. *Allium sativum*. I. Universidade Federal de Viçosa.
Departamento de Informática. Programa de Pós-Graduação em
Estatística Aplicada e Biometria. II. Título.

CDD 22. ed. 519.536

GUILHERME ALVES PUIATTI

**TÉCNICA DE AGRUPAMENTO NA SELEÇÃO DE MODELOS DE REGRESSÃO
NÃO LINEARES PARA DESCRIÇÃO DO ACÚMULO DE MATÉRIA SECA EM
PLANTAS DE ALHO**

Dissertação apresentada à
Universidade Federal de Viçosa, como
parte das exigências do Programa de
Pós-Graduação em Estatística Aplicada
e Biometria, para obtenção do título de
Magister Scientiae.

APROVADA: 19 de fevereiro de 2014

Adésio Ferreira

Moisés Nascimento
(Coorientador)

Paulo Roberto Cecon
(Orientador)

*"A linguagem é apenas o instrumento da ciência,
e as palavras não passam de símbolos das ideias."*

Samuel Johnson

AGRADECIMENTOS

Agradeço a Deus, por me conceder forças e perseverança ao longo desta jornada.

A Universidade Federal de Viçosa, pela oportunidade deste mestrado.

A CAPES e ao Programa REUNI, pela concessão da bolsa de estudos.

A FAPEMIG, pelo financiamento do projeto de pesquisa.

Ao meu pai, Mário, meu maior exemplo de luta e dedicação, e minha mãe, Janice, que sempre me incentivou e acreditou em mim.

Aos meus irmãos, Gustavo e Vinícius, que estiveram ao meu lado sempre que precisei.

A minha namorada, Nádia, pela compreensão, apoio e otimismo.

Ao meu orientador e amigo, Paulo Roberto Cecon, os maiores e mais sinceros agradecimentos pela dedicação, paciência, confiança, e por todas suas palavras de sabedoria.

Aos Coorientadores, Moysés Nascimento, Ana Carolina Campana Nascimento e Fabyano Fonseca e Silva, pelos ensinamentos ao longo do mestrado, e ao Adésio Ferreira, membro da banca, pelas sugestões para este trabalho.

A todo o corpo docente do Departamento de Estatística da UFV, pelos conhecimentos transmitidos e pela vontade de ajudar.

Aos professores do Departamento de Fitotecnia da UFV, pela concessão dos dados do experimento.

Aos colegas do mestrado, pela convivência por estes dois anos, juntos dividimos nossas alegrias e aprendizados.

E a todos que contribuíram para a realização deste trabalho, muito obrigado.

SUMÁRIO

RESUMO.....	vi
ABSTRACT	viii
1 Introdução	1
2 Revisão Bibliográfica	3
2.1 A cultura do alho.....	3
2.2 Divergência genética	4
2.3 Medidas de dissimilaridade	5
2.4 Análise de agrupamento.....	6
2.5 Regressão não linear	7
2.6 Avaliadores de qualidade de ajuste para modelos de regressão	9
3 Materiais e Métodos	11
3.1 Descrição do experimento	11
3.2 Modelos de regressão não linear	12
3.3 Método dos mínimos quadrados ordinários.....	13
3.4 Processo iterativo de Gauss-Newton	15
3.5 Avaliadores de qualidade de ajuste.....	21
3.5.1 Coeficiente de determinação	21
3.5.2 Quadrado médio do resíduo	21
3.5.3 Desvio médio absoluto dos resíduos.....	22
3.5.4 Critérios de informação de Akaike e Bayesiano	22
3.6 Dissimilaridade entre modelos ajustados	23
3.7 Método UPGMA	24
3.8 Critério de Mojena para determinação do número de grupos	25
3.9 Aspectos computacionais.....	26
4 Resultados e Discussão	27
4.1 Ajuste dos modelos de regressão não lineares e avaliadores de qualidade de ajuste.....	27
4.2 Agrupamento dos modelos pelos avaliadores de qualidade de ajuste	36

4.3 Agrupamento dos modelos segundo as estimativas dos parâmetros	39
4.4 Comparação dos resultados dos agrupamentos	41
5 Conclusão	44
6 Referências Bibliográficas	45
APÊNDICE	52
A – Dendrogramas das análises de agrupamento.....	53
B – Script das análises no <i>software</i> R	59

RESUMO

PUIATTI, Guilherme Alves. M.Sc., Universidade Federal de Viçosa, fevereiro de 2014. **Técnica de agrupamento na seleção de modelos de regressão não lineares para descrição do acúmulo de matéria seca em plantas de alho.** Orientador: Paulo Roberto Cecon. Coorientadores: Moysés Nascimento, Fabyano Fonseca e Silva e Ana Carolina Campana Nascimento.

Estudos de divergência genética entre indivíduos ou populações de plantas e sua trajetória de crescimento são de grande importância em programas de melhoramento, sendo essenciais para a obtenção de informações relevantes para um manejo adequado das plantas. Das técnicas empregadas para tal, a análise de agrupamento e modelos de regressão são amplamente utilizados. Assim, o objetivo deste estudo foi identificar e agrupar modelos de regressão não linear que melhor se ajustam na descrição do acúmulo de matéria seca total da planta do alho ao longo do tempo (60, 90, 120 e 150 dias após plantio). Foram utilizados 15 acessos de alho pertencentes ao Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH/UFV). Os modelos de regressão não linear ajustados para cada um dos acessos foram: Brody, Gompertz, Logístico, Mitscherlich e von Bertalanffy. A qualidade de ajuste dos modelos foi determinada pelo coeficiente de determinação (R^2); quadrado médio do resíduo (QMR); desvio médio absoluto dos resíduos (DMA); critério de informação de Akaike (AIC); e critério de informação Bayesiano (BIC). Então, para cada acesso, os modelos foram submetidos a análise de agrupamento, com os avaliadores de qualidade de ajuste considerados como variáveis, utilizando o algoritmo UPGMA, a distância generalizada de Mahalanobis como medida de dissimilaridade, e número de grupos determinado pelo método de Mojena. Depois, os modelos ajustados para cada acesso foram novamente agrupados seguindo o mesmo critério, mas utilizando os parâmetros com interpretação biológica como variáveis, e os resultados dos diferentes agrupamentos foram então confrontados. Comparando os resultados dos agrupamentos, observou-se que os modelos Gompertz, Logístico, e von Bertalanffy apresentaram melhores resultados quanto aos avaliadores de qualidade de ajuste, e tiveram resultados próximos quanto a estes e quanto as estimativas dos parâmetros. Estes três modelos se mostraram

eficientes para descrição de matéria seca total da planta em acessos de alho, especialmente o modelo Logístico.

ABSTRACT

PUIATTI, Guilherme Alves. M.Sc., Universidade Federal de Viçosa, February, 2014. **Clustering technique for selection of nonlinear regression models for the description of dry matter accumulation of garlic plants.** Adviser: Paulo Roberto Cecon. Co-advisers: Moysés Nascimento, Fabyano Fonseca e Silva and Ana Carolina Campana Nascimento.

Studies of genetic divergence between individuals or populations of plants and its growth path are of great importance in improvement programs, being essential for an appropriate handling through relevant information. Cluster analysis and regression models are techniques largely employed for such. The objective of this study was to identify and to group nonlinear regression models that best describe the dry matter accumulation of garlic plants over time (60, 90, 120 and 150 days after planting). 15 garlic accessions belonging to the Vegetable Germplasm Bank of Universidade Federal de Viçosa (BGH/UFV) were selected. The following nonlinear regression models were adjusted for each one of the accesses: Brody, Gompertz, Logistic, Mitscherlich and von Bertalanffy. The fit quality of the models was measured by the determination coefficient (R^2); mean squared error (MSE); mean absolute deviation of the error (DMA); Akaike information criterion (AIC); and Bayesian information criterion (BIC). Then, for each access, the models were submitted the grouping analysis, with the appraisers of fit quality used as variables, using the UPGMA algorithm, the Mahalanobis distance as dissimilarity measure, and Mojena's method to obtain the number of groups. The adjusted models for each access were again grouped following the same criterion, but using the parameters with biological interpretation as variables, and the results of the different clustering analysis were then confronted. Comparing the results, it was observed that the Gompertz, Logistic and von Bertalanffy models presented better results, and they had close results in fit quality and estimates of the parameters. These three models, especially the Logistic, were efficient for the description of dry matter accumulation of garlic plants.

1 INTRODUÇÃO

O estudo da trajetória de crescimento das plantas é de grande importância para a realização de um manejo adequado para estas. Esse conhecimento auxilia na preparação de técnicas de cultivo, colheita, conservação e também para a detecção de problemas no desenvolvimento de culturas. Curvas de acúmulo de matéria seca e de nutrientes servem como um parâmetro para a indicação da demanda e dos cuidados em cada etapa do desenvolvimento da planta (PÔRTO et al., 2007).

O alho (*Allium sativum* L.), planta herbácea de propagação vegetativa, é de grande importância devido a seus valores econômicos, culinários, nutricionais e medicinais. As variedades de alho possuem um bom potencial para o comércio, em especial para a industrialização, o que tem grande influência no sistema de agricultura familiar. O alho se destaca quanto ao valor nutricional pelo seu conteúdo em calorias, proteínas, carboidratos, fósforo, tiamina (B1) e vitamina B6 (TRANI, 2013). Além disso, ele também demonstra várias aplicações na medicina, principalmente por causa das suas características imuno estimulantes, antiarterioscleróticas, anticancerígenas e antimicrobianas (MARCHIORI, 2013).

O Brasil apresenta diversas variedades na cultura de alho, que surgiram por mutações somáticas e seleções de características desejáveis para seu manejo na agricultura (SOUZA; MACÊDO, 2009; HONORATO, 2012). Visando auxiliar na preparação de técnicas de cultivo, colheita e conservação de diferentes cultivares, tornam-se necessárias pesquisas com a finalidade de estudar o desenvolvimento em grupos de alhos que apresentem características semelhantes.

A análise de crescimento descreve o desenvolvimento vegetal em função do tempo, podendo ser avaliado de diversas maneiras, gerando conhecimentos de valor prático e informações referentes ao crescimento e comportamento dos genótipos que podem ser utilizadas pelos produtores (SHARMA et al., 1993). A curva de acúmulo de matéria seca do vegetal é importante em especial pelas informações a respeito do manejo da adubação e de outros tratamentos, pois a taxa de crescimento da planta varia durante seu desenvolvimento e exige diferentes quantidades de nutrientes em cada fase.

A abordagem mais adequada para o estudo de crescimento em plantas é a utilização de modelos de regressão não lineares, nos quais as formulações de possíveis modelos são baseadas em considerações teóricas inerentes ao fenômeno que se deseja modelar (MAZUCHELI; ACHCAR, 2002). Características como a variação do peso da matéria seca da planta variam com taxas decrescentes ao longo do tempo, apresentando um comportamento não linear. Os modelos não lineares oferecem vantagens por possuírem parâmetros com interpretação biológica (MAIA et al., 2009). A evolução dos computadores e software ainda são um incentivo para esse estudo, tornando análises mais ágeis e precisas (SILVA; PADOVANI, 2006).

A escolha do melhor modelo para a situação, no entanto, pode ser uma tarefa difícil. Existem diversos avaliadores de qualidade de ajuste calculados de diferentes maneiras, que priorizam diferentes características em suas avaliações (SILVEIRA, 2011). A escolha do modelo deve contar não somente com resultados de avaliadores, mas também com o conhecimento a respeito do objeto em estudo. Existe ainda a necessidade de agrupar os modelos mais próximos para os dados em estudo. A técnica da análise de agrupamento fornece uma maneira de se obter grupos homogêneos, de forma a reunir os indivíduos em um determinado número de grupos para que exista homogeneidade dentro de cada grupo e heterogeneidade entre eles (JOHNSON; WICHERN, 1992). Dessa forma, espera-se que modelos dentro de um mesmo grupo sejam similares quanto à interpretação e ao desempenho.

Diante deste contexto, este trabalho teve como objetivo ajustar, agrupar e avaliar modelos de regressão não lineares utilizados para descrição do acúmulo de matéria seca em plantas de alho ao longo do tempo, comparando resultados de avaliadores de qualidade de ajuste e estimativas dos parâmetros.

2 REVISÃO BIBLIOGRÁFICA

2.1 A cultura do alho

O alho (*Allium sativum* L.) merece destaque quanto a seus valores econômicos, culinários, nutricionais e medicinais. Plantado no Brasil principalmente por pequenos agricultores, este é muito utilizado como tempero ou mesmo como ingrediente principal na culinária. Composto por calorias, proteínas, carboidratos, fósforo, tiamina (B1) e vitamina B6, o alho possui propriedades medicinais para ser utilizado como analgésico, anti-inflamatório, anti-séptico, antibacteriano, antimicótico, antiviral, anticoagulante, além de proteger o fígado, estimular o sistema imunológico, e até reduzir o colesterol (TRANI, 2013; MARCHIORI, 2013). Segundo dados do IBGE (2013), a produção brasileira de alho no ano de 2012 foi de 108.393 toneladas, havendo um decréscimo de 24,35% em relação ao ano anterior. Os Estados com a maior produção são Goiás, com uma participação de 32,98%, Santa Catarina com 17,82%, Minas Gerais com 16,73%, e Rio Grande do Sul com uma participação de 16,13%.

A planta do alho é herbácea, normalmente alcança 50 centímetros de altura ou mais, dependendo da cultivar plantada. Possui plantas lanceoladas, com limbo medindo de 20 a 30 centímetros. Seu pseudocaule é formado pelas bainhas das folhas as quais implantam-se em um caule pequeno e achatado. As gemas do caule desenvolvem-se formando cada uma um bulbilho, que em seu conjunto formam o bulbo. Uma importante característica que difere os cultivares é o número de bulbilhos que integram o bulbo, número que é muito variável, podendo ir de 5 a 56. Os bulbilhos são formados pela: folha de reserva; folha de brotação, caracterizada por uma gema capaz de originar uma nova planta; e folha de proteção, que possui coloração branca ou arroxeadada. Os bulbilhos em geral têm morfologia ovoide arqueada, falciforme, e estão ligados ao caule pela base, estando cobertos por várias folhas, que em conjunto constituem a capa (SOUZA; MACÊDO, 2009).

A cultura do alho exige baixas temperaturas para que ocorra a bulbificação, assim como fotoperíodos mais longos aceleram a formação de bulbos. Normalmente temperaturas médias de 12,8 a 23,9 °C favorecem o

desenvolvimento do alho, porém existem diferenças entre os cultivares quanto à resposta à temperatura e ao fotoperíodo com conseqüente variação de resposta destas às regiões de cultivo e épocas de plantio (MUELLER et al., 1990).

Vários estudos utilizando o alho foram realizados desde a década de 80, principalmente em relação aos seus efeitos na prevenção de doenças cardiovasculares tais como: redução das concentrações séricas de LDL, triglicerídeos, redução da pressão arterial, aumento da atividade fibrinolítica e inibição da agregação plaquetária. A ação antioxidante da aliina, alicina e do ajoeno justificam o efeito do alho sobre as LDL, pois inibem a peroxidação lipídica por meio da inibição da enzima xantina-oxidase e de eicosanoides. O alho também eleva a capacidade total antioxidante do organismo devido à ação dos bioflavonóides quercetina e campferol. O alho ainda possui importante ação antimicrobiana, inicialmente descrita por Pasteur. Em 1944 foi testada a ação bactericida da alicina, com efeitos positivos na inibição do crescimento de várias bactérias, tanto gram positivas quanto gram negativas (ANJO, 2004; MARCHIORI, 2013).

2.2 Divergência genética

Estudos sobre a divergência genética entre indivíduos ou populações nas espécies vegetais são de grande importância para programas de melhoramento envolvendo hibridações, pois fornecem parâmetros para a identificação de progenitores que possibilitam maior efeito heterótico na progênie, e maior probabilidade de obter genótipos superiores em gerações segregantes (VIDIGAL et al., 1997). No melhoramento genético, a diferença nas frequências dos alelos das populações, ou seja, a dissimilaridade genética, tem importância fundamental para a escolha de genótipos que serão utilizados como progenitores, já que a divergência genética entre os parentais é um indicativo da expressão heterótica nas progênies (FALCONER, 1981).

A dissimilaridade genética normalmente é calculada a partir de um conjunto de características, utilizando técnicas de estatística multivariada. Como se trata de uma análise unificadora que permite integrar as múltiplas informações das características extraídas das unidades experimentais, a estatística

multivariada proporciona maior oportunidade de escolha de parentais divergentes em programas de melhoramento (DAS; GUPTA, 1984; JOHNSON; WICHERN, 1992).

Existem duas maneiras de inferir sobre diversidade genética, sendo uma de natureza quantitativa, e a outra de natureza preditiva (CRUZ et al., 2004). Os métodos preditivos de diversidade genética têm sido muito utilizados, principalmente pelo fato de que, sendo baseados em diferenças morfológicas, fisiológicas e moleculares dos genótipos, dispensam a obtenção das combinações híbridas entre eles, o que é vantajoso especialmente quando o número de genitores cuja diversidade se deseja conhecer é elevado (CARVALHO et al., 2003). Por esses métodos as informações múltiplas de cada cultivar são expressas em medidas de dissimilaridade, que representam a diversidade existente no conjunto de acessos estudados.

2.3 Medidas de dissimilaridade

Em análises de agrupamento, é necessária a definição de uma medida de distância, através da qual são definidos critérios para avaliar a proximidade de dois elementos, e se estes podem fazer parte de um mesmo grupo ou não (AAKER et al., 2001). A medida de dissimilaridade representa o quão distante dois objetos comparados estão: quanto maior este valor, mais diferentes eles serão. Um grande número de distâncias tem sido propostas e utilizadas em análise de agrupamento, e cada uma delas produz um determinado tipo de agrupamento (MINGOTI, 2005).

As medidas de dissimilaridade mais utilizadas em análise de agrupamento, segundo Cruz et al. (2004), são:

- Distância Euclidiana: definida como a soma da raiz quadrada da diferença entre dois elementos em suas respectivas dimensões. Atributos com maiores variâncias tendem a se destacar sobre os demais.
- Distância Euclidiana Quadrática: a distância é definida como o somatório dos quadrados das diferenças entre os valores dos elementos para todas as variáveis. Esta medida aumenta o peso

dos objetos mais distantes, ressaltando a diferença entre os grupos. Eficiente em implementações computacionais.

- Distância Euclidiana Ponderada: esta distância propõe diferentes pesos para variáveis que o pesquisador julgar mais importante para definir semelhança.
- Distância Generalizada de Mahalanobis: esta medida leva em consideração a existência de correlações entre os caracteres analisados por meio da matriz de variâncias e covariâncias residuais, porém necessita de experimentos com repetições (CARGNELUTTI FILHO et al., 2008).

2.4 Análise de agrupamento

A análise de agrupamento é composta por um conjunto de técnicas estatísticas cujo propósito é classificar os dados, unindo-os pelas semelhanças ou pelas diferenças, de acordo com o contexto. Os elementos de um mesmo grupo devem ser o mais semelhante possível entre si, enquanto a diferença entre os grupos deve ser a maior possível (FERNAU; SAMSON, 1990). A distância entre indivíduos é normalmente determinada pela distância euclidiana ou pelo coeficiente de correlação, muito comumente utilizados em estudos geofísicos (GONG; RICHMAN, 1995). Os métodos de agrupamento mais utilizados são os hierárquicos e os de otimização.

Segundo Bertan et al. (2006), nos métodos hierárquicos os indivíduos são agrupados por um processo que se repete em vários níveis, estabelecendo um dendrograma, sem preocupação com o número ótimo de grupos. Dentre os principais métodos de agrupamento hierárquicos então: o método do vizinho mais próximo; o método do vizinho mais distante; o método o UPGMA (*Unweighted Pair Group Method with Arithmetic Mean*); e o método da variância mínima de Ward. Estes métodos utilizam apenas a matriz de distâncias, dispensando a necessidade de recorrer aos dados originais. Já nos métodos de otimização, os grupos são estabelecidos aperfeiçoando determinado critério de agrupamento, onde os grupos formados são mutuamente exclusivos, diferente dos métodos

hierárquicos (CRUZ et al., 2004). O método de Tocher e o método de Tocher modificado são exemplos de métodos de otimização.

O método da ligação média entre grupos (UPGMA) tem sido muito utilizado para estudos de diversidade genética, tendo como vantagem sobre os demais métodos o fato de considerar médias aritméticas das medidas de dissimilaridade, evitando caracterizar a dissimilaridade por valores extremos entre os indivíduos considerados (CRUZ et al., 2004), como ocorre com os métodos do vizinho mais próximo e do vizinho mais distante. Outra vantagem deste método é que ele produz menor distorção na representação das similaridades entre indivíduos de um dendrograma (SOKAL; ROHLF, 1962). O método UPGMA tem sido utilizado em estudos de divergência genética de diversas culturas, tais como gergelim (ARRIEL et al., 2006), goiabeiras (GOMES FILHO et al., 2010), e pimentas (FARIA et al. 2012).

2.5 Regressão não linear

Avaliar a relação entre uma variável dependente com uma ou mais variáveis independentes é uma das tarefas mais comuns em análises estatísticas, e este objetivo pode ser alcançado através dos modelos de regressão, os quais se dividem em duas classes distintas: os modelos lineares e os modelos não lineares (MAZUCHELI; ACHCAR, 2002). Modelos de regressão lineares são modelos matemáticos que relacionam o comportamento de uma variável dependente com uma (ou mais, no caso de regressão multivariada) variável independente, e é expressa como função linear dos coeficientes de regressão. No entanto, em muitas situações práticas algumas de suas suposições, tais como a normalidade e a linearidade nos parâmetros, não são satisfeitas. Este fato alavancou o desenvolvimento de novas técnicas estatísticas para os modelos de regressão, surgindo então outra classe de modelos, os modelos de regressão não lineares. Um modelo é classificado como não linear se pelo menos uma das derivadas parciais da função esperança em relação ao parâmetro é função de parâmetros desconhecidos (PRUDENTE, 2009).

Um modelo de regressão não linear pode ser classificado como “intrinsecamente linear” se ele pode ser reduzido a um modelo linear através de

uma reparametrização apropriada. O termo “intrinsecamente linear” também pode ser usado para referir-se a modelos que podem ser linearizados via alguma transformação. Na prática, modelos não lineares são reduzidos com o objetivo de obter estimativas para os parâmetros. O inconveniente de uma transformação é que o parâmetro pode perder sua interpretação intrínseca, assim como é possível alterar a estrutura e distribuição do erro. Ou seja, se os erros do modelo original satisfizerem as suposições usuais de normalidade, independência e homogeneidade de variâncias, os erros do novo modelo na maioria dos casos não irão satisfazer tais suposições. Caso não seja possível obter uma reparametrização ou uma transformação apropriada que reduza o modelo à forma linear então o modelo é chamado de “intrinsecamente não linear” (MAZUCHELI; ACHCAR, 2002).

Ainda, segundo Mazucheli e Achcar (2002), os modelos de regressão não linear são da seguinte forma:

$$y = f(x; \theta) + \varepsilon$$

em que $y = (y_1, y_2, \dots, y_n)^t$ é o vetor de variáveis resposta, $x = (x_1, x_2, \dots, x_n)^t$ é o vetor de variáveis explicativas, $\theta = (\theta_1, \theta_2, \dots, \theta_p)^t$ é o vetor de parâmetros, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^t$ representa o vetor de erros aleatórios, e $f(x; \theta) = (f(x_1; \theta), f(x_2; \theta), \dots, f(x_n; \theta))^t$ é uma função das variáveis regressoras e dos parâmetros chamada de função esperança ou função de regressão (sendo n o número de observações e p o número de parâmetros).

Nos modelos lineares, utilizando-se o Método dos Mínimos Quadrados Ordinários, é possível ajustar um modelo que estime os parâmetros minimizando a soma de quadrados dos erros. No entanto, quando aplicado em regressão não linear, este método pode trazer equações de solução muito difícil (RATKOWSKY, 1983). Assim, os métodos iterativos oferecem uma alternativa para a obtenção das estimativas, obtendo soluções melhores a cada iteração a partir de uma aproximação inicial. Dentre estes destacam-se o método de Gauss-Newton ou método da linearização; o método Steepest-Descent ou método do gradiente; e o método de Marquardt (BATES; WATTS, 1988).

A utilização de funções matemáticas não lineares, desenvolvidas empiricamente para relacionar peso-idade, tem se mostrado adequada para descrição de curvas de crescimento, já que essas funções sintetizam um grande número de medidas em apenas alguns parâmetros contendo significado biológico, facilitando a interpretação e o entendimento do fenômeno (BROWN et al., 1976; DENISE; BRINKS, 1985). Aplicações de modelos de regressão não linear são comuns em Biologia, quando plantas e organismos crescem com o tempo ou a idade, mas existem também aplicações em Economia e Engenharia para esses tipos de modelos (REGAZZI, 2003).

Os modelos mais usados para descrição de crescimento de plantas e animais são: Brody, Gompertz, Logístico, Richards, Mitscherlich, Weibull, e von Bertalanffy (REIS, 2012). Sarmiento et al. (2006) utilizaram os modelos Brody, Logístico, Gompertz, Von Bertalanffy e Richards para descrição da curva de crescimento de ovinos Santa Inês. Maia et al. (2009), em um estudo de comparação de modelos de regressão não linear em bananeiras, fizeram uso dos modelos Brody, Gompertz, Logístico e von Bertalanffy. Reis (2012) utilizou os modelos Mitscherlich, Gompertz, Logístico, Brody, von Bertalanffy, e Meloun I e II para ajuste de modelos de regressão não linear para descrição de crescimento em grupos de alho.

2.6 Avaliadores de qualidade de ajuste para modelos de regressão

Segundo Silveira et al. (2011), avaliadores de qualidade de ajuste são ferramentas estatísticas que permitem comparar diferentes modelos e encontrar o mais indicado para os dados. Na literatura existem diversas propostas de medidas para verificar a qualidade de modelos ajustados. Em geral, de acordo com Sousa (2012), o modelo desejável é o mais parcimonioso, ou seja, aquele que envolve o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta.

Dentre os avaliadores mais utilizados estão o coeficiente de determinação ajustado (R_{aj}^2), e o desvio padrão residual. Existem dúvidas a respeito do uso do coeficiente de determinação (R^2) para regressão não linear, pois este não possui

significado óbvio neste caso (RATKOWSKY, 1983). Por outro lado, Souza (1998) afirma que a qualidade do ajustamento pode ser medida pelo quadrado do coeficiente de correlação entre os valores observados e preditos, mesmo no caso de modelos de regressão não linear, como nos trabalhos de Reis (2012) e Maia et al. (2009).

Sarmiento et al. (2006), Maia et al. (2009) e Reis (2012) utilizam o quadrado médio do resíduo (*QMR*) e o desvio médio absoluto dos resíduos (*DMA*). Os critérios de informação de Akaike e Bayesiano são outras estatísticas muito utilizadas como medidas de comparação da qualidade de ajuste de modelos baseados no máximo da função de verossimilhança, e dependem do número de observações e parâmetros do modelo em estudo. Estes critérios foram utilizados por Sousa (2012), Silva et al. (2011), e Silveira et al. (2011).

No entanto, existem dúvidas a respeito da qualidade do ajuste de diferentes modelos devido ao grande número de avaliadores existentes na literatura. Dessa forma, uma análise considerando todos os avaliadores de qualidade simultaneamente torna-se necessária para identificar o modelo que melhor se ajusta aos dados de crescimento. Uma proposta para tal método seria a utilização de métodos de classificação multivariados, como a análise de agrupamento. Esta análise forma grupos homogêneos de acordo com a avaliação simultânea de várias características de interesse (SILVA et al., 2011).

3 MATERIAIS E MÉTODOS

3.1 Descrição do experimento

O experimento foi realizado em área experimental pertencente ao setor de olericultura do Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV), no município de Viçosa, Zona da Mata de Minas Gerais, nas coordenadas geográficas: 20º 45' de latitude sul e 42º 51' de longitude oeste, com altitude média de 650 m, no período de março a novembro de 2010.

O experimento avaliou 15 acessos de alho, registrados no Banco de Germoplasma de Hortaliças (BGH/UFV). O delineamento experimental utilizado foi o de blocos completos casualizados, com quatro repetições.

As plantas colhidas foram submetidas ao processo de cura de campo e de galpão, por 3 e 60 dias, respectivamente, expostas ao sol e em galpão. Após a cura de campo, as plantas foram submetidas à toaleta cortando a parte aérea a 1,0 cm acima dos bulbos e retirando-se as raízes. A avaliação de matéria seca das partes das plantas desses acessos foi realizada em quatro períodos sendo: o primeiro período e inicial (60 dias após plantio); segundo período (90 dias após plantio); terceiro período (120 dias após plantio); e o quarto período e final (150 dias após plantio).

A avaliação foi feita sobre a matéria seca total da planta (MSTP), que foi obtida com o somatório da matéria seca da folha (MSF), do pseudocaule (MSP), do bulbo (MSB) e da raiz (MSR). As MSF, MSP, MSB e MSR foram obtidas por secagem em estufa com ventilação forçada a 65°C, até massa constante, expressa em gramas por planta.

Na Tabela 1 são apresentados os 15 acessos utilizados neste trabalho, com seus respectivos nomes comuns e origem.

Tabela 1. Relação dos 15 acessos de alho registrados no Banco de Germoplasmas de Hortaliças da Universidade Federal de Viçosa (BGH/UFV)

Nº	Acesso	Nome comum/variedade	Origem
1	4484	Juiz de Fora	Florestal – MG
2	4493	Cateto Roxo	Florestal – MG
3	4567	Cateto Roxo	Gouvéia – MG
4	4814	–	Santa Tereza – MG
5	4818	–	Santa Tereza – MG
6	4820	–	Santa Tereza – MG
7	4822	–	Santa Leopoldina – ES
8	4830	Sapé	Rio Possmoser – ES
9	4839	–	Cedrolândia, Nova Venecia – ES
10	4844	–	Afonso Cláudio – ES
11	5938	Cateto Roxo	Viçosa – MG
12	5939	Amarante Aimorés	Viçosa – MG
13	5963	Roxo Capim Branco	–
14	5965	Vila Verde	–
15	6396	–	–

3.2 Modelos de regressão não linear

Para cada um dos 15 acessos, foram ajustados os cinco modelos apresentados na Tabela 2. A variável independente representa os dias após plantio (DAP), e a variável dependente foi a média da matéria seca total da planta (MSTP), em gramas, para cada período.

Tabela 2. Modelos de regressão não linear utilizados para descrever o acúmulo de matéria seca

Modelo	Função	Referência
A	$y_i = \beta_1(1 - \beta_2 e^{(-\beta_3 x_i)}) + \varepsilon_i$	Brody (1945)
B	$y_i = \beta_1 e^{(-\beta_2 e^{(-\beta_3 x_i)})} + \varepsilon_i$	Gompertz (GOMPertz; PHILos, 1825)
C	$y_i = \frac{\beta_1}{1 + \beta_2 e^{(-\beta_3 x_i)}} + \varepsilon_i$	Logístico (RATKowsky, 1983)
D	$y_i = \beta_1(1 - e^{(\beta_3 \beta_2 - \beta_3 x_i)}) + \varepsilon_i$	Mitscherlich (1919)
E	$y_i = \beta_1(1 - \beta_2 e^{(-\beta_3 x_i)})^3 + \varepsilon_i$	von Bertalanffy (1957)

Para cada modelo, temos que:

- β_1 é o parâmetro que representa o peso assintótico do acesso;
- β_2 é um parâmetro de locação, sem interpretação biológica;
- β_3 é o parâmetro da taxa de maturidade dos acessos;
- y_i é a observação da variável resposta: a matéria seca total da planta (MSTP);
- x_i é a variável preditora, representada pelos períodos da avaliação de matéria seca da planta (dias após plantio, DAP);
- ε_i é o termo de erro aleatório, supondo erros independentes com média zero e variância constante, tal que $\varepsilon_i \sim N(0, \sigma^2)$.

3.3 Método dos mínimos quadrados ordinários

Segundo Mazucheli e Achcar (2002), o método de estimação por mínimos quadrados é usado para análise de dados em que as observações são constituídas por variáveis dependentes y_i obtidas em diferentes níveis da variável independente x_i , ($i = 1, 2, \dots, n$). Assume-se que a relação entre variável resposta e variável independente pode ser adequadamente representada por uma equação da forma:

$$y = f(x; \theta) + \varepsilon$$

onde temos que $y = (y_1, y_2, \dots, y_n)^t$ é o vetor de variáveis resposta (dependente), $x = (x_1, x_2, \dots, x_n)^t$ é o vetor de explicativas (independente), $\theta = (\theta_1, \theta_2, \dots, \theta_p)^t$ é o vetor de parâmetros, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^t$ é o vetor de erros aleatórios, e $f(x; \theta) = (f(x_1; \theta), f(x_2; \theta), \dots, f(x_n; \theta))^t$ é a função das variáveis regressoras e dos parâmetros chamada de função esperança ou função de regressão.

Como a função esperança é contínua e admite derivadas de primeira e segunda ordens com relação aos parâmetros, define-se a soma de quadrados dos erros por:

$$S(\theta) = \sum_{i=1}^n [y_i - f(x_i; \theta)]^2$$

em notação vetorial:

$$S(\theta) = [y - f(\theta)]^t [y - f(\theta)] = \|y - f(\theta)\|^2$$

A seguir, $F(\theta)$ é a matriz Jacobiana da função vetor $f(\theta) = (f_1(\theta), f_2(\theta), \dots, f_n(\theta))^t$. A função $f(\theta)$ tem dimensão $n \times 1$ e o vetor de parâmetros θ é de dimensão $p \times 1$:

$$F(\theta) = \frac{\partial f(\theta)}{\partial \theta^t} = \begin{bmatrix} \frac{\partial f_1(\theta)}{\partial \theta_1} & \frac{\partial f_1(\theta)}{\partial \theta_2} & \dots & \frac{\partial f_1(\theta)}{\partial \theta_p} \\ \frac{\partial f_2(\theta)}{\partial \theta_1} & \frac{\partial f_2(\theta)}{\partial \theta_2} & \dots & \frac{\partial f_2(\theta)}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(\theta)}{\partial \theta_1} & \frac{\partial f_n(\theta)}{\partial \theta_2} & \dots & \frac{\partial f_n(\theta)}{\partial \theta_p} \end{bmatrix}$$

O estimador de mínimos quadrados $\hat{\theta}$ segue então a equação:

$$\left[\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right]_{\boldsymbol{\theta}' = \hat{\boldsymbol{\theta}}'} = \mathbf{0}$$

$$\begin{aligned} \frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} &= \frac{\partial}{\partial \boldsymbol{\theta}'} [\mathbf{y} - f(\boldsymbol{\theta})]' [\mathbf{y} - f(\boldsymbol{\theta})] \\ &= -2[\mathbf{y} - f(\boldsymbol{\theta})]' \mathbf{F}(\boldsymbol{\theta}) \\ &= -2\mathbf{F}(\boldsymbol{\theta})' [\mathbf{y} - f(\boldsymbol{\theta})] \end{aligned}$$

De tal forma:

$$\mathbf{F}(\hat{\boldsymbol{\theta}})' = [\mathbf{y} - f(\hat{\boldsymbol{\theta}})] = \mathbf{0}$$

Sendo que o sistema equações normais para a regressão não linear é:

$$\begin{bmatrix} \frac{\partial f_1(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} & \frac{\partial f_2(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} & \dots & \frac{\partial f_n(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} \\ \frac{\partial f_1(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_2} & \frac{\partial f_2(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_2} & \dots & \frac{\partial f_n(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_p} & \frac{\partial f_2(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_p} & \dots & \frac{\partial f_n(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_p} \end{bmatrix} \begin{pmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} f_1(\hat{\boldsymbol{\theta}}) \\ f_2(\hat{\boldsymbol{\theta}}) \\ \vdots \\ f_n(\hat{\boldsymbol{\theta}}) \end{bmatrix} \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Estas equações são não lineares em relação aos estimadores dos parâmetros. As soluções destas equações podem ser extremamente difíceis de serem obtidas, e na maioria dos casos um método iterativo de resolução de equações normais não lineares deve ser utilizado.

3.4 Processo iterativo de Gauss-Newton

O algoritmo de Gauss-Newton é um dos métodos mais utilizados para se obter estimativas de mínimos quadrados não lineares. Supondo $\hat{\boldsymbol{\theta}}_0$ como uma aproximação inicial para $\hat{\boldsymbol{\theta}}$, temos as expansões de Taylor das componentes de $f(\boldsymbol{\theta})$ na vizinhança de $\hat{\boldsymbol{\theta}}_0$:

$$f(x_i, \theta) = f(x_i, \hat{\theta}_0) + \frac{\partial f(x_i, \hat{\theta}_0)}{\partial \theta^t} (\theta - \hat{\theta}_0)$$

que produzem a aproximação linear:

$$f(\theta) \cong f(\hat{\theta}_0) + F(\hat{\theta}_0)(\theta - \hat{\theta}_0)$$

onde:

$$F(\hat{\theta}_0) = \left[\frac{\partial f(\theta)}{\partial \theta^t} \right]_{\theta = \hat{\theta}_0^t}$$

O problema de minimizar a função de soma dos quadrados dos erros aleatórios torna-se minimizar a função $S(\theta)$ associada à aproximação linear, assim:

$$S(\theta) = \|y - f(\hat{\theta}_0) - F(\hat{\theta}_0)(\theta - \hat{\theta}_0)\|^2$$

Considerando $y - f(\hat{\theta}_0) = E_0$ e $\theta - \hat{\theta}_0 = \Delta\hat{\theta}_0$, a partir do resultado anterior temos que

$$\begin{aligned} S(\theta) &= \|E_0 - F(\hat{\theta}_0)\Delta\hat{\theta}_0\|^2 \\ &= [E_0 - F(\hat{\theta}_0)\Delta\hat{\theta}_0]^t [E_0 - F(\hat{\theta}_0)\Delta\hat{\theta}_0] \\ &= E_0^t E_0 - E_0^t F(\hat{\theta}_0)\Delta\hat{\theta}_0 - \Delta\hat{\theta}_0^t F(\hat{\theta}_0)^t E_0 + \Delta\hat{\theta}_0^t F(\hat{\theta}_0)^t F(\hat{\theta}_0)\Delta\hat{\theta}_0 \end{aligned}$$

Derivando $S(\theta)$ em relação a $\Delta\hat{\theta}_0$:

$$\begin{aligned}\frac{\partial S(\theta)}{\partial \Delta \hat{\theta}_0} &= \mathbf{0} - 2\mathbf{F}(\hat{\theta}_0)' \mathbf{E}_0 + 2\mathbf{F}(\hat{\theta}_0)' \mathbf{F}(\hat{\theta}_0) \Delta \hat{\theta}_0 \\ &= 2 \left[\mathbf{F}(\hat{\theta}_0)' \mathbf{F}(\hat{\theta}_0) \Delta \hat{\theta}_0 - \mathbf{F}(\hat{\theta}_0)' \mathbf{E}_0 \right]\end{aligned}$$

Igualando a derivada a zero, teremos o Sistema de Equações Normais:

$$\mathbf{F}(\hat{\theta}_0)' \mathbf{F}(\hat{\theta}_0) \Delta \hat{\theta}_0 = \mathbf{F}(\hat{\theta}_0)' \mathbf{E}_0$$

no qual \mathbf{E}_0 desempenha o papel de variável dependente \mathbf{y} e $\mathbf{F}(\hat{\theta}_0)'$ o da matriz \mathbf{X} dos modelos lineares. Se $\mathbf{F}(\hat{\theta}_0)$ apresenta posto coluna completo, o valor de $\Delta \hat{\theta}_0$ que minimiza $S(\theta)$ será:

$$\Delta \hat{\theta}_0 = \left[\mathbf{F}(\hat{\theta}_0)' \mathbf{F}(\hat{\theta}_0) \right]^{-1} \mathbf{F}(\hat{\theta}_0)' \mathbf{E}_0$$

Pode-se definir um vetor $\hat{\theta}_1 = \hat{\theta}_0 + \Delta \hat{\theta}_0$ de estimativas atualizadas de θ . A partir deste procedimento, pode-se obter utilizando $\hat{\theta}_1$ outro conjunto de estimativas atualizadas $\hat{\theta}_2$, e assim por diante. Uma k -ésima iteração será:

$$\begin{aligned}\hat{\theta}_{k+1} &= \hat{\theta}_k + \Delta \hat{\theta}_k \\ &= \hat{\theta}_k + \left[\mathbf{F}(\hat{\theta}_k)' \mathbf{F}(\hat{\theta}_k) \right]^{-1} \mathbf{F}(\hat{\theta}_k)' \mathbf{E}_k\end{aligned}$$

em que

$$\begin{aligned}\mathbf{F}(\hat{\theta}_k) &= \left[\frac{\partial f(\theta)}{\partial \theta'} \right]_{\theta=\hat{\theta}_k} \\ \mathbf{E}_k &= \mathbf{y} - \mathbf{f}(\hat{\theta}_k) \\ \hat{\theta}_k &= \left[\hat{\theta}_{1k} \quad \hat{\theta}_{2k} \quad \dots \quad \hat{\theta}_{pk} \right]'\end{aligned}$$

O processo continua até a convergência, que ocorre quando $|\hat{\theta}_{k+1} - \hat{\theta}_k| < \delta$, em que δ é algum valor fixo bem pequeno. O valor de $S(\hat{\theta}_k)$ pode ser avaliado, para garantir a redução deste valor durante dada iteração. Caso $S(\hat{\theta}_{k+1})$ seja maior que $S(\hat{\theta}_k)$, usa-se $\Delta\hat{\theta}_k/2$ como o vetor de incremento.

Uma característica interessante deste método ocorre quando a função esperança é linear. Nesse caso, o processo converge em uma única iteração, partindo-se de qualquer valor inicial (MAZUCHELI; ACHCAR, 2002).

Exemplo: Considerando os dados do acesso 15 (BGH 6396), a seguir é demonstrado o ajuste do modelo Logístico utilizando o método de Gauss-Newton, e com os valores iniciais dos parâmetros $\hat{\theta}_{1,0} = 22$, $\hat{\theta}_{2,0} = 4000$ e $\hat{\theta}_{3,0} = 0,07$. Os valores iniciais podem ser obtidos de diferentes formas, tais como interpretação física dos parâmetros, ou uma regressão linear simples a partir de uma transformação para linearização da função. Os vetores dos dias após plantio, matéria seca total da planta (em gramas) e dos valores iniciais são:

$$\mathbf{x} = \begin{bmatrix} 60 \\ 90 \\ 120 \\ 150 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0,8223 \\ 4,4625 \\ 16,8500 \\ 22,7750 \end{bmatrix} \text{ e } \hat{\boldsymbol{\theta}} = \begin{bmatrix} 22 \\ 4000 \\ 0,07 \end{bmatrix}$$

Para estes valores temos que:

$$S(\hat{\boldsymbol{\theta}}_0) = \sum_{i=1}^n [y_i - f(\mathbf{x}_i; \boldsymbol{\theta})]^2 = 40,0466$$

A partir do modelo Logístico, são calculadas suas derivadas em relação a cada parâmetro:

$$f(\mathbf{x}) = \frac{\theta_1}{1 + \theta_2 e^{(-\theta_3 x)}}$$

$$\frac{\partial f(x, \theta_1, \theta_2, \theta_3)}{\partial \theta_1} = \frac{1}{1 + \theta_2 e^{(-\theta_3 x)}}$$

$$\frac{\partial f(x, \theta_1, \theta_2, \theta_3)}{\partial \theta_2} = \frac{\theta_1 e^{(-\theta_3 x)}}{(1 + \theta_2 e^{(-\theta_3 x)})^2}$$

$$\frac{\partial f(x, \theta_1, \theta_2, \theta_3)}{\partial \theta_3} = \frac{\theta_1 \theta_2 e^{(-\theta_3 x)}}{(1 + \theta_2 e^{(-\theta_3 x)})^2}$$

A partir da primeira observação $x_1 = 60$, temos que:

$$F_{1;1} = \frac{1}{1 + \theta_2 e^{(-\theta_3 x_1)}} \Bigg|_{\theta_1=22; \theta_2=4000; \theta_3=0,07} = \frac{1}{1 + 4000 e^{(-0,07 \cdot 60)}} = 0,0164$$

$$F_{1;2} = \frac{\theta_1 e^{(-\theta_3 x)} }{(1 + \theta_2 e^{(-\theta_3 x)})^2} \Bigg|_{\theta_1=22; \theta_2=4000; \theta_3=0,07} = \frac{22 \cdot e^{(-0,07 \cdot 60)}}{(1 + 4000 \cdot e^{(-0,07 \cdot 60)})^2} = -0,0001$$

$$F_{1;3} = \frac{\theta_1 \theta_2 e^{(-\theta_3 x)} }{(1 + \theta_2 e^{(-\theta_3 x)})^2} \Bigg|_{\theta_1=22; \theta_2=4000; \theta_3=0,07} = \frac{22 \cdot 4000 \cdot e^{(-0,07 \cdot 60)}}{(1 + 4000 \cdot e^{(-0,07 \cdot 60)})^2} = 21,2907$$

E o mesmo cálculo é realizado para as demais observações, com os resultados apresentados na Tabela 3.

Tabela 3 Dados, estimativas, resíduo e derivadas para os dados, considerando a aproximação inicial $\hat{\theta}_0 = [22 \quad 4000 \quad 0,07]^T$

x_i	y_i	$f(x_i, \hat{\theta}_0)$	E_0	$F_{i;1}$	$F_{i;2}$	$F_{i;3}$
60	1,0646	0,3607	0,4615	0,0164	-0,0001	21,2907
90	4,0500	2,6362	1,8263	0,1198	-0,0006	208,8307
120	15,6375	11,5821	5,2678	0,5265	-0,0014	658,1513
150	24,0750	19,8172	2,9578	0,9008	-0,0005	294,9319

O vetor de incremento agora é calculado utilizando a matriz $F(\hat{\theta}_0)$ das derivadas $F_{i,j}$:

$$\Delta\hat{\theta}_0 = \left[F(\hat{\theta}_0)' F(\hat{\theta}_0) \right]^{-1} F(\hat{\theta}_0)' E_0 = \begin{bmatrix} 1,2818 \\ -1978,9137 \\ 0,0028 \end{bmatrix}$$

Assim, a estimativa atualizada $\hat{\theta}_1$ é calculada:

$$\hat{\theta}_1 = \hat{\theta}_0 + \Delta\hat{\theta}_0 = \begin{bmatrix} 22 \\ 4000 \\ 0,07 \end{bmatrix} + \begin{bmatrix} 1,2818 \\ -1978,9137 \\ 0,0028 \end{bmatrix} = \begin{bmatrix} 23,2818 \\ 2021,0863 \\ 0,0728 \end{bmatrix}$$

O vetor $\hat{\theta}_1$ é considerado a estimativa atualizada de θ , uma vez que a soma de quadrados residual é $S(\hat{\theta}_1) = 3,0286$, menor que $S(\hat{\theta}_0) = 40,0466$. Agora, uma nova iteração é efetuada com $\hat{\theta}_1$, e assim até o modelo convergir. A Tabela 4 apresenta as estimativas atualizadas a cada iteração:

Tabela 4 Estimativas dos parâmetros e soma de quadrados residuais nas iterações do algoritmo de Gauss-Newton para ajuste do modelo

k	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$S(\hat{\theta}_k)$
0	22	4000	0,07	40,0466
1	23,2818	2021,0863	0,0728	3,0286
2	23,8109	3160,7522	0,0759	0,6400
3	23,7253	4161,7897	0,0772	0,1655
4	23,7052	4516,6029	0,0776	0,0920
5	23,7002	4571,8668	0,0777	0,0914
6	23,6998	4575,0990	0,0777	0,0914
7	23,6998	4575,2397	0,0777	0,0914

3.5 Avaliadores de qualidade de ajuste

Para a verificação da qualidade do ajuste dos modelos, foram calculados os seguintes avaliadores: coeficiente de determinação (R^2); quadrado médio do resíduo (QMR); desvio médio absoluto dos resíduos (DMA); critério de informação de Akaike (AIC); e critério de informação Bayesiano (BIC).

3.5.1 Coeficiente de determinação

O coeficiente de determinação (R^2), segundo Kennedy (2008), deve representar a proporção da variação da variável dependente que é explicada pela variação da variável independente. Seu valor é calculado por:

$$R^2 = 1 - \frac{SQR}{SQT}$$

no qual:

- SQR é a soma de quadrados dos resíduos, calculada com base na diferença dos valores estimados e observados da matéria seca total da planta de alho;
- SQT é a soma de quadrados total, com correção pela média, calculada a partir da diferença entre o valor observado e a média das observações da matéria seca total.

Quanto maior o coeficiente de determinação, melhor a qualidade do modelo ajustado.

3.5.2 Quadrado médio do resíduo

O quadrado médio do resíduo (QMR) é dado por:

$$QMR = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n - p}$$

em que:

- n é o número de observações;
- p é o número de parâmetros;
- y_i é o valor observado da matéria seca total da observação i ;
- \hat{y}_i é o valor estimado da matéria seca total da observação i para o modelo analisado.

Quanto menor for o valor do quadrado médio do resíduo, melhor será o modelo ajustado.

3.5.3 Desvio médio absoluto dos resíduos

O desvio médio absoluto dos resíduos (DMA), proposto por Sarmento et al. (2006), é calculado como o somatório dos desvios entre os valores observados e estimados, dividido pelo número de observações:

$$DMA = \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{n}$$

em que:

- n é o número de observações;
- y_i é o valor observado da matéria seca total da observação i ;
- \hat{y}_i é o valor estimado da matéria seca total da observação i para o modelo analisado.

Quanto menor o desvio médio absoluto dos resíduos, melhor o modelo ajustado.

3.5.4 Critérios de informação de Akaike e Bayesiano

O critério de informação de Akaike (AIC), proposto por Akaike (1974), e o critério de informação Bayesiano (BIC), por Schwarz (1978), são estatísticas para

comparação da qualidade de ajuste do modelo baseados no máximo da função de verossimilhança, e dependem do número de observações e parâmetros do modelo em questão. O critério de informação de Akaike admite a existência de um modelo real, mas desconhecido, que descreve os dados. Ele tenta então escolher dentre um grupo de modelos avaliados aquele que minimiza a divergência de Kull-Leibler, que está relacionada à informação perdida por se usar o modelo aproximado ao invés do “real”. Já o critério de informação Bayesiano maximiza a probabilidade de se identificar o verdadeiro modelo dentre os avaliados (SOUSA, 2012). O valor do *AIC* é calculado por:

$$AIC = 2k - 2\log L(\hat{\theta})$$

E a estimativa do *BIC* é dada por:

$$BIC = k \log(n) - 2\log L(\hat{\theta})$$

em que:

- n é o número de observações;
- k é o número de parâmetros no modelo;
- $L(\hat{\theta})$ é o máximo da função de verossimilhança.

O melhor modelo, segundo os critérios de informação de Akaike e Bayesiano, é aquele que apresenta menor valor.

3.6 Dissimilaridade entre modelos ajustados

As medidas de dissimilaridade são a base para a análise de agrupamento. A medida utilizada foi a distância euclidiana quadrática, também conhecida por distância de Karl Pearson, pois as características avaliadas como variáveis são calculadas de diferentes maneiras, apresentando diferentes escalas, e estas variáveis não apresentam correlação. A distância euclidiana quadrática entre dois grupos é obtida por:

$$d_{rs}^2 = \sum_{j=1}^p (y_{rj} - y_{sj})^2$$

em que:

- d_{rs}^2 representa a distância euclidiana quadrática entre os modelos r e s ;
- y_{rj} é a característica(avaliador) do indivíduo(modelo) r ;
- y_{sj} representa a característica(avaliador) do indivíduo(modelo) s ;
- p é a dimensão(número de variáveis) da amostra;

3.7 Método UPGMA

O método UPGMA (*Unweighted Pair-Group Method using Arithmetic Averages*), ou método da ligação média entre grupos não ponderado, foi o método de agrupamento utilizado para a análise. Para cada acesso, os modelos ajustados foram agrupados segundo os avaliadores de qualidade de ajuste e os parâmetros estimados.

Este método de agrupamento utiliza as médias aritméticas (não ponderadas) das medidas de dissimilaridade, atribuindo sempre o mesmo peso aos elementos que estão sendo integrados. A construção dos dendrogramas é estabelecida pelos modelos ajustados de menor dissimilaridade. A distância entre um modelo k e um grupo, formado pelos modelos i e j , é dada por:

$$d_{(ij)k} = \text{média}(d_{ik}; d_{jk}) = \frac{d_{ik} + d_{jk}}{2}$$

ou seja, $d_{(ij)k}$ é dada pela média do conjunto das distâncias dos pares de indivíduos (i e k) e (j e k).

Segundo Ferreira (2008), o método engloba todos os outros métodos hierárquicos aglomerativos, sendo necessária apenas a determinação adequada dos parâmetros (α , β e γ) da fórmula geral proposta para atualizar a distância

entre as entidades i e j , recém-agrupadas, e a próxima entidade k . A fórmula de atualização é:

$$d_{(ij)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma |d_{ik} - d_{jk}|$$

na qual, para o método UPGMA.

$$\alpha_i = \frac{n_i}{n_i + n_j}, \alpha_j = \frac{n_j}{n_i + n_j}, \beta = 0 \text{ e } \gamma = 0.$$

Assim, d representa a distância entre as entidades recém-agrupadas i e j e a entidade k ; n_i , n_j e n_k representam o número de indivíduos de cada entidade. De forma que essa nova expressão se torne idêntica a aquela apresentada anteriormente.

3.8 Critério de Mojena para determinação do número de grupos

Os métodos hierárquicos geram dendrogramas, e para determinar o ponto de corte destes e assim definir o número de grupos, utiliza-se um procedimento baseado no tamanho relativo dos níveis de fusões (distâncias) no dendrograma (MOJENA, 1977). A proposta é selecionar o número de grupos no passo j que, primeiramente, satisfizer a seguinte inequação:

$$\alpha_j > \theta_k$$

onde α_j é o valor de distância do nível de fusão correspondente aos passos j ($j=1, 2, \dots, g-1$), e θ_k é o valor referencial de corte, dado por:

$$\theta_k = \bar{\alpha} + k \hat{\sigma}_\alpha$$

onde $\bar{\alpha}$ e $\hat{\sigma}_\alpha$ são respectivamente a média e o desvio padrão dos valores de α ; k é uma constante cujo valor adotado foi de $k = 1,25$, conforme sugerido por Milligan e Cooper (1985), como regra de parada na definição do número de grupos. Assim:

$$\bar{\alpha} = \frac{1}{g-1} \sum_{j=1}^{g-1} \alpha_j \quad \text{e} \quad \hat{\sigma}_\alpha = \sqrt{\frac{\sum_{j=1}^{g-1} \alpha_j^2 - \frac{1}{g-1} \left(\sum_{j=1}^{g-1} \alpha_j \right)^2}{g-2}}$$

3.9 Aspectos computacionais

O ajuste dos modelos de regressão não linear descritos na Tabela 2, os cálculos dos avaliadores de qualidade de ajuste, e as análises de agrupamento foram obtidas no software livre R versão 2.14.2 (R DEVELOPMENT CORE TEAM, 2011).

4 RESULTADOS E DISCUSSÃO

4.1 Ajuste dos modelos de regressão não lineares e avaliadores de qualidade de ajuste

Os 15 acessos de alho, apresentados na Tabela 1 do item 3.1, foram submetidos ao ajuste dos cinco modelos de regressão não linear contidos na Tabela 2, item 3.2. Para todos os acessos houve convergência dos modelos. São apresentadas nas Tabelas 5 a 19 as estimativas dos parâmetros, assim como os avaliadores de qualidade de ajuste, para cada acesso avaliado.

Observando-se os avaliadores de qualidade de ajuste, pode-se perceber que os resultados foram concordantes na maioria dos casos: os modelos com melhores resultados para coeficiente de determinação (R^2) também obtiveram, em geral, melhores resultados para o quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA) e para os critérios de informação de Akaike (AIC) e Bayesiano (BIC). O acesso 13 obteve melhores resultados em média para R^2 , QMR e DMA (Tabela 17), enquanto o acesso 14 teve os menores valores para AIC e BIC (Tabela 18). O modelo Logístico obteve as melhores médias para os resultados dos avaliadores de qualidade de ajuste, seguido pelos modelos Gompertz, von Bertalanffy, Mitscherlich e Brody, respectivamente (Tabela 20).

As estimativas dos parâmetros foram bem variadas dentre os diferentes modelos (Tabela 21). Em média, o peso assintótico (referente ao parâmetro β_1) foi muito alto nos modelos Brody e Mitscherlich, enquanto os modelos Logístico, Gompertz e von Bertalanffy apresentaram um valor menor e mais próximo. As estimativas de $\hat{\beta}_2$ apresentaram a maior variação dentre os modelos, chegando a 251.433,8212 no caso do modelo Logístico. Mas como este não possui interpretação biológica, isso não representa problema. A estimativa do parâmetro β_3 , que representa a velocidade do crescimento do acesso, foi pequena para todos os modelos, sendo consideravelmente menor nos modelos Brody e Mitscherlich.

Tabela 5. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 1

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,5963	136,9670	5,5260	33,4853	31,0305	3027,0000	1,0000	0,0001
B – Gompertz	0,9971	0,9918	0,3060	13,7734	11,3185	29,9977	53,2521	0,0366
C – Logístico	0,9994	0,2095	0,1639	7,5539	5,0991	26,2500	2345,0000	0,0678
D – Mitscherlich	0,8954	35,4830	2,6524	28,0825	25,6277	5626,0000	58,7300	0,0001
E – von Bertalanffy	0,9963	1,2565	0,3739	14,7196	12,2648	33,4084	5,2868	0,0263

Tabela 6. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 2

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,8657	46,7045	3,0558	29,1817	26,7269	143,4000	1,1260	0,0020
B – Gompertz	0,9952	1,6584	0,3220	15,8298	13,3750	21,5300	8,5119	0,1196
C – Logístico	0,9956	1,5435	0,3466	15,5425	13,0877	21,5600	3563000,0000	0,1498
D – Mitscherlich	0,8657	46,7045	3,0558	29,1817	26,7269	143,4000	60,5400	0,0020
E – von Bertalanffy	0,9450	19,1375	2,1217	25,6129	23,1581	25,5244	13,9268	0,0400

Tabela 7. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 3

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,8925	55,1131	3,3197	29,8439	27,3891	214,7731	1,1017	0,0016
B – Gompertz	0,9968	1,6474	0,3209	15,8031	13,3482	26,7941	9713,3660	0,0960
C – Logístico	0,9974	1,3262	0,3387	14,9355	12,4807	26,7712	35853,1400	0,1255
D – Mitscherlich	0,8925	55,1131	3,3197	29,8439	27,3891	214,7685	60,8720	0,0016
E – von Bertalanffy	0,9622	19,3685	2,1371	25,6609	23,2061	31,3325	13,5685	0,0396

Tabela 8. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 4

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,9199	17,3953	1,8646	25,2311	22,7763	86,7997	1,1785	0,0027
B – Gompertz	0,9966	0,7432	0,2155	12,6193	10,1644	18,1369	2526,4840	0,0827
C – Logístico	0,9980	0,4250	0,2164	10,3833	7,9285	18,1088	46990,3785	0,1063
D – Mitscherlich	0,9199	17,3953	1,8646	25,2311	22,7763	86,7968	59,9800	0,0027
E – von Bertalanffy	0,9766	5,0906	1,1039	20,3159	17,8611	20,5922	14,1255	0,0406

Tabela 9. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 5

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,8515	53,1404	3,1932	29,6981	27,2433	13821,3242	1,0007	0,0001
B – Gompertz	0,9939	2,1858	0,3698	16,9343	14,4795	24,2706	778,4158	0,0680
C – Logístico	0,9961	1,4081	0,3841	15,1753	12,7205	23,9633	23225,9386	0,0956
D – Mitscherlich	0,9169	29,7301	2,2941	27,3750	24,9202	14678,6721	58,6204	0,0001
E – von Bertalanffy	0,9784	7,7188	1,3366	21,9810	19,5261	28,0233	10,5615	0,0362

Tabela 10. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 6

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,8435	35,5318	2,5860	28,0880	25,6332	13136,8590	1,0006	0,0001
B – Gompertz	0,9959	0,9230	0,2408	13,4860	11,0312	19,5357	516,0921	0,0637
C – Logístico	0,9980	0,4510	0,2244	10,6211	8,1663	19,2031	14321,5602	0,0909
D – Mitscherlich	0,9189	18,4042	1,8802	25,4566	23,0018	13663,8800	59,1810	0,0001
E – von Bertalanffy	0,9849	3,4237	0,8900	18,7292	16,2744	22,2501	10,6220	0,0363

Tabela 11. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 7

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,8374	36,5844	2,7251	28,2048	25,7500	10228,3033	1,0007	0,0001
B – Gompertz	0,9935	1,4578	0,3239	15,3140	12,8592	20,9574	162,5506	0,0516
C – Logístico	0,9980	0,4607	0,2527	10,7062	8,2513	20,2783	3133,8415	0,0756
D – Mitscherlich	0,9488	11,5135	1,5008	23,5804	21,1256	9743,6272	57,7909	0,0001
E – von Bertalanffy	0,9898	2,2858	0,6482	17,1132	14,6584	23,0778	9,1910	0,0347

Tabela 12. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 8

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,7226	48,9248	3,1722	29,3675	26,9126	7181,1285	1,0007	0,0001
B – Gompertz	0,9966	0,5964	0,1934	11,7389	9,2841	17,2034	504,5555	0,0627
C – Logístico	0,9982	0,3175	0,1812	9,2173	6,7625	16,8230	19528,7395	0,0926
D – Mitscherlich	0,9245	13,3082	1,5825	24,1599	21,7050	8280,3663	59,4529	0,0001
E – von Bertalanffy	0,9850	2,6398	0,7814	17,6892	15,2344	19,8505	9,9008	0,0350

Tabela 13. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 9

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,7478	128,5406	5,1840	33,2313	30,7765	12923,2254	1,0007	0,0001
B – Gompertz	0,9975	1,2589	0,2810	14,7274	12,2726	29,1920	477,9992	0,0620
C – Logístico	0,9988	0,5889	0,2463	11,6884	9,2336	28,4976	19542,1384	0,0925
D – Mitscherlich	0,9271	37,1309	2,6388	28,2641	25,8093	14910,2765	59,8396	0,0001
E – von Bertalanffy	0,9867	6,7732	1,2545	21,4582	19,0034	33,6070	9,9787	0,0350

Tabela 14. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 10

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,8588	27,7941	2,5086	27,1056	24,6508	7504,1919	1,0009	0,0001
B – Gompertz	0,9974	0,5139	0,1974	11,1436	8,6887	19,9853	116,1574	0,0470
C – Logístico	0,9996	0,0749	0,0985	3,4377	0,9829	18,8892	3447,3483	0,0749
D – Mitscherlich	0,9486	10,1255	1,4689	23,0666	20,6117	7582,0640	59,0829	0,0001
E – von Bertalanffy	0,9953	0,9284	0,4159	13,5094	11,0545	22,0008	8,1470	0,0326

Tabela 15. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 11

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,8455	67,5844	3,6606	30,6598	28,2050	17106,1415	1,0007	0,0001
B – Gompertz	0,9990	0,4242	0,1641	10,3759	7,9211	27,2536	334,4316	0,0586
C – Logístico	0,9999	0,0551	0,0780	2,2109	-0,2439	26,4732	12288,7979	0,0883
D – Mitscherlich	0,9416	25,5519	2,1511	26,7692	24,3144	16909,0205	60,6872	0,0001
E – von Bertalanffy	0,9921	3,4535	0,8989	18,7639	16,3091	30,7239	10,4504	0,0357

Tabela 16. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 12

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,9269	16,1839	1,7987	24,9424	22,4876	101,6490	1,1465	0,0023
B – Gompertz	0,9936	1,4120	0,2971	15,1864	12,7316	18,8686	1392,2886	0,0766
C – Logístico	0,9965	0,7764	0,3113	12,7942	10,3393	18,8594	18402,5275	0,0970
D – Mitscherlich	0,9269	16,1839	1,7987	24,9424	22,4876	101,6547	58,5202	0,0023
E – von Bertalanffy	0,9779	4,8963	1,0740	20,1603	17,7054	21,2285	13,4979	0,0403

Tabela 17. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 13

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,9588	12,5508	1,5843	23,9255	21,4707	1334,2490	1,0112	0,0002
B – Gompertz	0,9944	1,6938	0,3556	15,9142	13,4593	23,9602	186,0253	0,0543
C – Logístico	0,9989	0,3270	0,2231	9,3352	6,8804	23,4066	2779,9070	0,0764
D – Mitscherlich	0,9588	12,5508	1,5843	23,9255	21,4707	1329,5380	57,9371	0,0002
E – von Bertalanffy	0,9909	2,7792	0,7277	17,8949	15,4401	25,9162	10,9327	0,0376

Tabela 18. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 14

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,8325	53,2526	3,4033	29,7065	27,2517	5691,0000	1,0010	0,0001
B – Gompertz	0,9989	0,3409	0,1893	9,5018	7,0470	28,9521	47,5609	0,0358
C – Logístico	0,9999	0,0004	0,0076	-17,1201	-19,5750	25,2200	2073,0000	0,0671
D – Mitscherlich	0,9380	19,7106	1,9959	25,7310	23,2761	6539,4690	59,3143	0,0001
E – von Bertalanffy	0,9984	0,5019	0,2206	11,0489	8,5941	31,5712	5,3554	0,0267

Tabela 19. Coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e estimativas dos parâmetros β_1 , β_2 e β_3 segundo modelos estudados, para as médias da matéria seca total da planta (MSTP) do acesso 15

MODELO	Avaliadores de qualidade de ajuste					Estimativas dos parâmetros		
	R^2	QMR	DMA	AIC	BIC	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	0,8573	45,5162	3,1993	29,0786	26,6238	10290,0000	1,0010	0,0001
B – Gompertz	0,9979	0,6544	0,2152	12,1103	9,6555	24,9297	142,6805	0,0491
C – Logístico	0,9997	0,0915	0,1066	4,2386	1,7837	23,7000	4575,0000	0,0777
D – Mitscherlich	0,9470	16,9212	1,8783	25,1206	22,6658	10160,0000	59,4500	0,0001
E – von Bertalanffy	0,9952	1,5356	0,5620	15,5220	13,0671	27,6624	8,5199	0,0331

Tabela 20. Médias do coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC), e respectivas média, desvio padrão e coeficiente de variação ($CV\%$) para os modelos estudados

MODELO	Médias dos avaliadores de qualidade de ajuste				
	R^2	QMR	DMA	AIC	BIC
A – Brody	0,8371	52,1189	3,1187	28,7833	26,3285
B – Gompertz	0,9963	1,1001	0,2661	13,6306	11,1757
C – Logístico	0,9983	0,5370	0,2120	8,0480	5,5932
D – Mitscherlich	0,9247	24,3884	2,1111	26,0487	23,5939
E – von Bertalanffy	0,9836	5,4526	0,9698	18,6786	16,2238
Média	0,9480	16,7194	1,3355	19,0378	16,5830
Desvio padrão	0,0689	22,0494	1,2571	8,5772	8,5772
CV%	7,2647	131,8794	94,1249	45,0532	51,7226

Tabela 21. Médias das estimativas dos parâmetros β_1 , β_2 e β_3 para os modelos estudados, com média, desvio padrão e coeficiente de variação ($CV\%$).

MODELO	Médias das estimativas dos parâmetros		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Brody	6852,6696	1,0381	0,0007
B – Gompertz	23,4378	1130,6914	0,0643
C – Logístico	22,5336	251433,8212	0,0919
D – Mitscherlich	7331,3022	59,3332	0,0007
E – von Bertalanffy	26,4513	10,2710	0,0353
Média	2851,2789	50527,0310	0,0386
Desvio padrão	3874,9152	112311,3354	0,0400
CV%	135,9010	222,2797	103,5598

4.2 Agrupamento dos modelos pelos avaliadores de qualidade de ajuste

Para cada acesso, os modelos foram agrupados através do algoritmo UPGMA considerando os avaliadores de qualidade de ajuste como variáveis, a

fim de identificar os modelos mais próximos quanto ao seu desempenho no ajuste dos dados. Foi utilizado o método de Mojena (1977) com $k = 1,25$ para determinar o ponto de corte nos dendrogramas gerados.

O número ótimo de grupos encontrado para todos os acessos avaliados foi de dois, exceto para o acesso 14, que obteve apenas um grupo. Na Tabela 22 são apresentados os grupos, com seus respectivos modelos e médias dos avaliadores de qualidade de ajuste. Em oito dos grupos divididos, temos os modelos Brody e Mitscherlich no Grupo 1, e os modelos Logístico, Gompertz e von Bertalanffy no Grupo 2. Os modelos com melhores médias para os avaliadores pertencem ao Grupo 2 (maiores valores de R^2 , menores valores para os demais avaliadores), indicando que os dados se ajustaram bem a estes modelos. Para os demais acessos, o modelo Brody foi o único no Grupo 1, enquanto os outros modelos ficaram no Grupo 2. Este modelo apresenta resultados mais distantes, principalmente quanto aos valores do quadrado médio do resíduo (QMR) que foram altos.

O coeficiente de determinação (R^2) variou de 0,5963 a 0,9588 nos modelos do Grupo 1, e no Grupo 2 variou de 0,972 a 0,9974. Os valores do quadrado médio do resíduo (QMR) variam no Grupo 1 de 12,5508 até 136,967, e no Grupo 2 de 0,5057 a 11,438. Para o desvio médio absoluto dos resíduos (DMA) a variação foi de 1,1633 a 5,526, e de 0,2373 a 1,1052, nos Grupos 1 e 2 respectivamente. Para os critérios de informação de Akaike e Bayesiano, o Grupo 1 apresentou variação de $11,7736 < AIC < 33,4853$ e $9,3188 < BIC < 31,0305$, enquanto no Grupo 2 temos que $9,3635 < AIC < 19,0345$ e $6,9087 < BIC < 16,5797$. Assim, o Grupo 2 demonstra melhor qualidade de ajuste nos seus modelos.

Tabela 22. Grupos com seus respectivos modelos e médias dos avaliadores de qualidade de ajuste para cada acesso considerado

	Acesso	Grupo 1	Grupo 2	Acesso	Grupo 1	Grupo 2	Acesso	Grupo 1	Grupo 2	Acesso	Grupo 1	Grupo 2
Modelos		A	B, C, D, E		A, D	B, C, E		A	B, C, D, E		A, D	B, C, E
R^2		0,5963	0,9720		0,8842	0,9895		0,7478	0,9776		0,9588	0,9948
QMR	1	136,9670	9,4852	5	41,4352	3,7709	9	128,5406	11,4380	13	12,5508	1,6000
DMA		5,5260	0,8740		2,7436	0,6968		5,1840	1,1052		1,5843	0,4355
AIC		33,4853	16,0323		28,5365	18,0302		33,2313	19,0345		23,9255	14,3814
BIC		31,0305	13,5775		26,0817	15,5754		30,7765	16,5797		21,4707	11,9266
Modelos		A, D	B, C, E		A, D	B, C, E		A, D	B, C, E		A, B, C, D, E	
R^2		0,8657	0,9786		0,8812	0,9930		0,9037	0,9974		0,9536	
QMR	2	46,7045	7,4465	6	26,9680	1,5992	10	18,9598	0,5057	14	14,7613	
DMA		3,0558	0,9301		2,2331	0,4518		1,9888	0,2373		1,1633	
AIC		29,1817	18,9951		26,7723	14,2788		25,0861	9,3635		11,7736	
BIC		26,7269	16,5403		24,3175	11,8239		22,6313	6,9087		9,3188	
Modelos		A, D	B, C, E		A	B, C, D, E		A	B, C, D, E		A	B, C, D, E
R^2		0,8925	0,9855		0,8374	0,9825		0,8455	0,9831		0,8573	0,9850
QMR	3	55,1131	7,4473	7	36,5844	3,9294	11	67,5844	7,3712	15	45,5162	4,8007
DMA		3,3197	0,9322		2,7251	0,6814		3,6606	0,8231		3,1993	0,6905
AIC		29,8439	18,7998		28,2048	16,6785		30,6598	14,5300		29,0786	14,2479
BIC		27,3891	16,3450		25,7500	14,2236		28,2050	12,0752		26,6238	11,7930
Modelos		A, D	B, C, E		A	B, C, D, E		A, D	B, C, E			
R^2		0,9199	0,9904		0,7226	0,9761		0,9269	0,9893			
QMR	4	17,3953	2,0862	8	48,9248	4,2155	12	16,1839	2,3616			
DMA		1,8646	0,5119		3,1722	0,6846		1,7987	0,5608			
AIC		25,2311	14,4395		29,3675	15,7013		24,9424	16,0469			
BIC		22,7763	11,9846		26,9126	13,2465		22,4876	13,5921			

A: Brody; B: Gompertz; C: Logístico; D: Mitscherlich; E: von Bertalanffy.

4.3 Agrupamento dos modelos segundo as estimativas dos parâmetros

Com o método da ligação média entre grupos (UPGMA), os modelos ajustados foram agrupados para os 15 acessos novamente, desta vez utilizando as estimativas dos parâmetros como variáveis. O ponto de corte em cada dendrograma foi estabelecido pelo método de Mojena (1977) para $k=1,25$. Apenas os parâmetros com interpretação biológica β_1 e β_3 foram considerados nesta análise, uma vez que o objetivo é avaliar os modelos quanto à qualidade com que descrevem o crescimento dos acessos. O parâmetro de β_2 possui apenas função de locação e não foi levado em consideração para o agrupamento e interpretação, pois poderia interferir na formação dos grupos devido a grandes variações nos seus valores.

As análises de agrupamento apresentaram, para todos os acessos, a mesma formação de grupos: o Grupo 1, com os modelos Brody e Mitscherlich; e o Grupo 2, formado pelos modelos Logístico, Gompertz e von Bertalanffy.

As médias das estimativas dos parâmetros para os agrupamentos são apresentadas na Tabela 23. O valor do peso assintótico no Grupo 1, representado pelas estimativas $\hat{\beta}_1$, tem grande variação entre os diferentes acessos, indo de 86,7982 g até 17.007,58 g. Estes valores, além da grande variação, são altos demais para representar o peso assintótico da matéria seca total da planta. Ainda para o Grupo 1, os valores para taxa de crescimento, os valores estimados $\hat{\beta}_3$, são consideravelmente pequenos: o maior é de 0,0027; mas a maioria das estimativas tem o valor mínimo de 0,0001. Essas estimativas tornam a curva de crescimento quase uma reta, não se mostrando boas representantes para essa descrição. As estimativas dos parâmetros para estes dois modelos são muito próximas, exceto para o parâmetro β_2 , e isso pode ser devido ao fato destes modelos serem apenas reparametrizações diferentes de um mesmo modelo.

Tabela 23. Grupos com seus respectivos modelos e médias das estimativas dos parâmetros $\hat{\beta}_1$ e $\hat{\beta}_3$ para cada acesso considerado

Acesso		Grupo 1	Grupo 2	Acesso		Grupo 1	Grupo 2
1	Modelos	A, D	B, C, E	9	Modelos	A, D	B, C, E
	$\hat{\beta}_1$	4326,5	29,8854		$\hat{\beta}_1$	13916,75	30,4322
	$\hat{\beta}_3$	0,0001	0,0436		$\hat{\beta}_3$	0,0001	0,0632
2	Modelos	A, D	B, C, E	10	Modelos	A, D	B, C, E
	$\hat{\beta}_1$	143,4	22,8715		$\hat{\beta}_1$	7543,128	20,2917
	$\hat{\beta}_3$	0,002	0,1031		$\hat{\beta}_3$	0,0001	0,0515
3	Modelos	A, D	B, C, E	11	Modelos	A, D	B, C, E
	$\hat{\beta}_1$	214,7708	28,2993		$\hat{\beta}_1$	17007,58	28,1502
	$\hat{\beta}_3$	0,0016	0,087		$\hat{\beta}_3$	0,0001	0,0609
4	Modelos	A, D	B, C, E	12	Modelos	A, D	B, C, E
	$\hat{\beta}_1$	86,7982	18,946		$\hat{\beta}_1$	101,6519	19,6522
	$\hat{\beta}_3$	0,0027	0,0765		$\hat{\beta}_3$	0,0023	0,0713
5	Modelos	A, D	B, C, E	13	Modelos	A, D	B, C, E
	$\hat{\beta}_1$	14250	25,419		$\hat{\beta}_1$	1331,894	24,4276
	$\hat{\beta}_3$	0,0001	0,0666		$\hat{\beta}_3$	0,0002	0,0561
6	Modelos	A, D	B, C, E	14	Modelos	A, D	B, C, E
	$\hat{\beta}_1$	13400,37	20,3296		$\hat{\beta}_1$	6115,235	28,5811
	$\hat{\beta}_3$	0,0001	0,0636		$\hat{\beta}_3$	0,0001	0,0432
7	Modelos	A, D	B, C, E	15	Modelos	A, D	B, C, E
	$\hat{\beta}_1$	9985,965	21,4378		$\hat{\beta}_1$	10225	25,4307
	$\hat{\beta}_3$	0,0001	0,054		$\hat{\beta}_3$	0,0001	0,0533
8	Modelos	A, D	B, C, E				
	$\hat{\beta}_1$	7730,747	17,959				
	$\hat{\beta}_3$	0,0001	0,0634				

A: Brody; B: Gompertz; C: Logístico; D: Mitscherlich; E: von Bertalanffy.

Para o Grupo 2, as estimativas do peso assintótico ($\hat{\beta}_1$) apresentaram uma dispersão bem menor que no outro grupo, com os valores mínimo e máximo de 17,959 g e 30,4322 g respectivamente. As estimativas para taxa de maturidade ($\hat{\beta}_3$) vão de 0,0432 a 0,1031. Tais estimativas dos parâmetros são mais apropriadas segundo a interpretação prática. Isso, somado ao fato dos avaliadores de qualidade de ajuste dos três modelos do Grupo 2 apresentarem melhores resultados, demonstra que estes modelos são mais indicados para a

descrição do acúmulo de matéria seca total da planta para acessos de alho. As estimativas dos parâmetros destes modelos foram bem próximas em geral. O modelo von Bertalanffy apresentou maiores estimativas para o parâmetro β_1 , e menores para β_2 e β_3 . Os modelos Gompertz e Logístico possuem estimativas bem próximas de β_1 , sendo que as estimativas de β_3 do Logístico foram as maiores. E as estimativas de β_2 para o modelo Logístico foram bem grandes, e apresentaram forte variação, chegando a 3.563.000; mas isto não indicou nenhum problema de ajuste, pois este é apenas um parâmetro de locação, como já discutido.

4.4 Comparação dos resultados dos agrupamentos

A Tabela 24 apresenta os resultados dos agrupamentos, com os modelos separados segundo avaliadores de qualidade de ajuste e parâmetros estimados.

Tabela 24. Resultados dos agrupamentos dos 15 acessos de alho, utilizando avaliadores de qualidade de ajuste e estimativas dos parâmetros com interpretação biológica, respectivamente

Acesso	Avaliadores		Parâmetros	
	Grupo 1	Grupo 2	Grupo 1	Grupo 2
1	A	B, C, D, E	A, D	B, C, E
2	A, D	B, C, E	A, D	B, C, E
3	A, D	B, C, E	A, D	B, C, E
4	A, D	B, C, E	A, D	B, C, E
5	A, D	B, C, E	A, D	B, C, E
6	A, D	B, C, E	A, D	B, C, E
7	A	B, C, D, E	A, D	B, C, E
8	A	B, C, D, E	A, D	B, C, E
9	A	B, C, D, E	A, D	B, C, E
10	A, D	B, C, E	A, D	B, C, E
11	A	B, C, D, E	A, D	B, C, E
12	A, D	B, C, E	A, D	B, C, E
13	A, D	B, C, E	A, D	B, C, E
14	A, B, C, D, E	-	A, D	B, C, E
15	A	B, C, D, E	A, D	B, C, E

A: Brody; B: Gompertz; C: Logístico; D: Mitscherlich; E: von Bertalanffy.

Enquanto o agrupamento por estimativas dos parâmetros foi homogêneo para todos os acessos, apresentando dois grupos compostos pelos mesmos modelos, o agrupamento por avaliadores de qualidade de ajuste teve diferentes resultados para distintos acessos. Para a maioria dos acessos a análise de agrupamento segundo os avaliadores foi concordante com o agrupamento pelas estimativas de parâmetros. Em oito destes, os modelos Brody e Mitscherlich formam um grupo, enquanto o outro é composto pelos modelos Logístico, Gompertz e von Bertalanffy. Para outros seis acessos, o modelo Brody diferiu dos demais, ficando isolado em um grupo. Apenas no acesso 14 não houve formação de grupos de modelos.

Para todos os acessos, nota-se a proximidade dos valores estimados $\hat{\beta}_1$ e $\hat{\beta}_3$ para os modelos de um mesmo grupo, assim como a diferença entre os valores de grupos distintos. Isso indica que, para interpretação física destes valores, os modelos do mesmo grupo são igualmente representativos. Já para os avaliadores de qualidade de ajuste, a diferença nem sempre é clara. Em geral, o modelo Brody teve as piores avaliações, seguido pelo modelo Mitscherlich. Por isso, o modelo de Brody acabou isolado em muitas das análises. Este modelo acaba mostrando pior desempenho na descrição de crescimento de plantas, quando comparado a outros modelos de regressão não linear, conforme resultados obtidos por Maia et al. (2009) na descrição de curvas de crescimento em bananeiras. Isso se dá, provavelmente, pelo fato deste modelo não apresentar comportamento sigmoideal, e não possuir ponto de inflexão, sendo um modelo de crescimento restrito (BRODY, 1945), e mais característico para curvas de crescimento de animais que de plantas. O modelo de Mitscherlich, apesar das estimativas de parâmetros muito próximas do modelo Brody, obteve valores melhores quanto aos avaliadores, mas estes não foram melhores que os resultados dos outros três modelos.

Os modelos Logístico, Gompertz e von Bertalanffy foram melhores segundo os avaliadores, e estiveram no mesmo grupo para todos os acessos. Na maioria das análises de agrupamento, os modelos Logístico e Gompertz foram os mais similares, sendo os melhores na avaliação da qualidade de ajuste. Diversos trabalhos chegam a uma conclusão favorável a estes modelos para o estudo de crescimento, principalmente para plantas. Pôrto et al. (2007) realizam ajustes de

boa qualidade utilizando o modelo Logístico para descrever acúmulos de matéria seca e cultura de cebola. Maia et al. (2009) concluem que o modelo Logístico foi o melhor ajustado para descrever curvas de crescimento em bananeiras. Fernandes (2012) encontra bons resultados dos modelos Logístico e Gompertz para curvas de crescimento do fruto do cafeeiro, sendo que o Gompertz indica melhores ajustes. Além disso, Reis (2012) também identificou que o modelo Logístico apresentou os melhores ajustes para descrever o acúmulo de matéria seca do bulbo, da raiz e total da planta para grupos de acessos de alho.

5 CONCLUSÃO

A análise de agrupamento para os avaliadores de qualidade de ajuste revelou uma maior diferença entre os ajustes do modelo Brody para os demais, e maior proximidade entre os modelos Logístico, Gompertz e von Bertalanffy. Já segundo a análise de agrupamento para as estimativas dos parâmetros com interpretação biológica ($\hat{\beta}_1$ e $\hat{\beta}_3$), os modelos Logístico, Gompertz e von Bertalanffy apresentaram resultados satisfatórios, com estimativas similares entre si e viáveis para interpretação na prática.

Para os 15 acessos de alho pertencentes ao Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa que foram utilizados neste trabalho, o modelo Logístico forneceu os melhores resultados referentes aos avaliadores de qualidade de ajuste na descrição do acúmulo de matéria seca total da planta.

6 REFERÊNCIAS BIBLIOGRÁFICAS

AAKER, D. A.; KUMAR, V.; DAY, G. S. **Pesquisa de marketing**, São Paulo: Atlas, 2001. 745p.

AKAIKE, H. A. New Look at the Statistical Model Identification. **IEEE Transactions on automatic control**, Minato-ku, Tokyo, Japan. v.19, n.6, p.716-723, Dec. 1974.

ANJO, D. F. C. Alimentos funcionais em angiologia e cirurgia vascular. **J. Vasc. Br.**, v. 3, n. 2, p. 145-154, 2004.

ARRIEL, N. H. C.; DI MAURO, A. O.; DI MAURO, S. M. Z.; BAKKE, O. A.; UNÊDA-TREVISOLI, S. H.; COSTA, M. M.; CAPELOTO, A.; CORRADO, A. R. Técnicas multivariadas na determinação da diversidade genética em gergelim usando marcadores RAPD. **Pesq. agropec. bras.**, Brasília, v. 41, n. 5, p. 801-809, 2006.

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

BERTAN, I; et al. Comparação de métodos de agrupamento na representação da distância morfológica entre genótipos de trigo. **R. Bras. Agrociência**, Pelotas, v. 12, n. 3, p. 279-286, jul-set, 2006.

BRODY, S. **Bioenergetics and Growth**. Rheinhold Publishing, New York. 1945

BROWN, J. E.; FITZHUGH JUNIOR, H. A.; CARTWRIGHT, T. C. A comparison of nonlinear models for describing weight-age relationships in cattle. **Journal of Animal Science**, Champaign, v. 42, n. 4, p. 810-818, Apr. 1976.

CARGNELUTTI FILHO, A.; RIBEIRO, N. D.; REIS, R. C. P. dos; SOUZA, J. R. de; JOST, E. Comparação de métodos de agrupamento para o estudo da divergência genética em cultivares de feijão. **Cienc. Rural, Santa Maria**, v. 38, n. 8, 2008.

CARVALHO, L. P.; LANZA, M. A.; FALIERI, J.; SANTOS, J. W. Análise da diversidade genética entre acessos do banco ativo de germoplasma de algodão. **Pesquisa Agropecuária Brasileira**, v.38, n.10, p.1149-1155, 2003.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos Biométricos Aplicados ao Melhoramento Genético**. v. 1, 3 ed., Viçosa: UFV, 480p. 2004.

DAS, P. K.; GUPTA, T. D. Multivariate analysis in black grain (*Vigna mungo* (L.) Herpper). **Indian Journal of Genetics and Plant Breeding**, New Delhi, 44(7):243-247, 1984.

DENISE, R. S. K.; BRINKS, J. S. Genetic and environmental aspects of the growth curve parameters in beef cows. **Journal of Animal Science**, Champaign, v. 61, n. 6, p. 1431-1440, 1985.

FALCONER, D. S. **Introduction to quantitative genetics**. 2.ed. London, Longman, 1981. 340p.

FARIA, P. N.; CECON, P. R.; SILVA, A. R.; FINGER, F. L.; SILVA, F. F. e; CRUZ, C. D.; SÁVIO, F. L. Métodos de agrupamento em estudo de divergência genética de pimentas. **Hortic. Bras.**, Vitoria da Conquista, v. 30, n. 3, p. 428-432. 2012.

FERNANDES, T. J. **Curva de crescimento do fruto do cafeeiro em diferentes alinhamentos de plantio utilizando modelos não lineares**. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras. Orientador: Joel Augusto Muniz, Lavras, 2012, 80f.

FERNAU, M. E.; SAMSON, P. J. Use of cluster analysis to define periods of similar meteorology and precipitation chemistry in Eastern North America. Part I:

Transport patterns. **Journal of Applied Meteorology**, Michigan, v.29, p.735-761, 1990.

FERREIRA, D. F. **Estatística Multivariada**. 1 ed. Lavras: Editora UFLA, 2008. 662p.

GOMES FILHO, A.; OLIVEIRA, J. G. de; VIANA, A. P.; SIQUEIRA, A. P. de O.; OLIVEIRA, M. G.; PEREIRA, M. G. Marcadores moleculares RAPD e descritores morfológicos na avaliação da diversidade genética de goiabeiras (*Psidium guajava* L.). **Acta Sci., Agron.**, Maringá, v. 32, n. 4, p. 627-633. 2010

GOMPERTZ, B.; PHILOS, T. Roy Soc London. 115-513. 1825.

GONG, X.; RICHMAN, M. B. On the application to growing season precipitation data in North America East of the rockies. **Journal of Climate**, Oklahoma, v. 8, p.897-931, 1995.

HONORATO, A. R. de F. **Avaliação de cultivares de alho na região de Mossoró-RN**. Dissertação (Mestrado em Fitotecnia) – Universidade Federal Rural do Semi-Árido. Orientadora: Maria Zuleide de Negreiros, Mossoró, 59f. 2012.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Levantamento Sistemático da Produção Agrícola. Rio de Janeiro, v.26 n.01 p.1-83 janeiro. 2013. Disponível em: <http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/lspa/lspa_201301.pdf>. Acesso em: 16/05/2013.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Jersey, USA: Englewood Cliffs, 642p. 1992.

KENNEDY, Peter. **A Guide to Econometrics**. Massachusetts: Blackwell Publishing, 2008.

MAIA, E.; SIQUEIRA, D. L.; SILVA, F. F.; PETERNELLI, L. A.; SALOMÃO, L. C. C. Método de comparação de modelos de regressão não-lineares em bananeiras. **Ciência Rural**, v.39, n.5, p.1380-1386, 2009.

MARCHIORI, V. F. **Propriedades funcionais do alho (*Allium Sativum L.*)**. 2005. Disponível em: <http://www.esalq.usp.br/siesalq/pm/alho_revisado.pdf>. Acesso em: 13/5/2013.

MAZUCHELI, J.; ACHCAR, J. A. Algumas considerações em regressão não linear. **Acta Scientiarum**, Maringá, v.24, n.6, p.1761-1770, 2002.

MILLIGAN, G. W.; COOPER, M. C. An examination of procedures for determining the number of cluster in a data set. **Psychometrika**, v. 50, p. 159-179, 1985.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 297p. 2005.

MITSCHERLICH, E. A. Das Gesetz des Pflanzenwachstums. **Landwirtsch Jahrb** 53: 167-182. 1919.

MOJENA, R. Hierárquical grouping method and stopping rules: an evaluation. **Computer Journal**, v. 20, p. 359-363, 1977.

MUELLER, S.; BIASI, J.; MENEZES SOBRINHO, J. A.; MULLER, J. J. V. Comportamento de cultivares de alho, plantio de junho. **Pesquisa Agropecuária Brasileira**, v.25, n.11, p.1561-1567. 1990.

PÔRTO, D. R. de Q.; CECILIO FILHO, A. B.; MAY, A.; VARGAS, P. F. Acúmulo de macronutrientes pela cultivar de cebola Superex estabelecida por semeadura direta. **Ciência Rural**, v.37, p.949-955, 2007.

PRUDENTE, A. A. **Modelos não-lineares de regressão: alguns aspectos de teoria assintótica**. Dissertação (Mestrado em Biometria e Estatística Aplicada) – Universidade Federal Rural de Pernambuco. Recife, 2009, 108f.

R DEVELOPMENT CORE TEAM (2011). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

RATKOWSKY, D. A. **Nonlinear regression modeling: a unified practical approach**. Marcel Dekker, New York. 1983.

REGAZZI, A. J. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não linear. **Revista Ceres**, Viçosa, v. 50, n. 287, p. 9-26, 2003.

REIS, R. M. dos. **Modelos de regressão não linear para descrição do crescimento de plantas de alho**. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa. Orientador: Paulo Roberto Cecon, Viçosa, 2012, 47f.

SARMENTO, J. L. R.; REGAZZI, A. J.; SOUZA, W. H.; TORRES, R. A.; BREDA, F.C.; MENEZES, G. R. O. Estudo da curva de crescimento de ovinos Santa Inês. **Revista Brasileira de Zootecnia**, v. 35, n. 2, p. 435-442, 2006.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, v. 6, p. 461-464, 1978.

SHARMA, B. D.; KAUL, H. N.; SINGH, M. Growth analysis of potato varieties in autumn in subtropical conditions. **New Botanist**, Lucknow, v. 20, n. 54, p. 55-64, 1993.

SILVA, N. A. M.; LANA, A. M. Q.; SILVA, F. F.; SILVEIRA, F. G.; BERGMANN, J. A. G.; SILVA, M. A.; TORAL, F. L. B. Seleção e classificação multivariada de

modelos de crescimento não lineares para bovinos Nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**. v. 63, n. 2, p. 364-371, 2011.

SILVA, N. R. da; PADOVANI, C. R. Utilização de componentes principais em experimentação agrônômica. **Energ. Agric.**, Botucatu, vol. 21, n.4, p.98-113, 2006.

SILVEIRA, F. G.; SILVA, F. F.; CARNEIRO, P. L. S.; MALHADO, C. H. M.; MUNIZ, J. A. Análise de agrupamento na seleção de modelos de regressão não-lineares para curvas de crescimento de ovinos cruzados. **Ciência Rural**, v.41, n.4, p.692-698, 2011

SOKAL, R. R.; ROHLF, F. J. The comparison of dendrograms by objective methods. **Taxon**, v.11, p.30-40, 1962.

SOUSA, I. F. **Ajuste de modelos não lineares na descrição de germinação de sementes de café (*coffea arábica* L.)**. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Lavras. Orientador: Joel Augusto Muniz, Lavras, 2012, 72f.

SOUZA, G. S. **Introdução aos modelos de regressão linear e não-linear**. 1 ed. Brasília: Embrapa, 489p. 1998.

SOUZA, R. J.; MACÊDO, F. S. **Cultura do alho: tecnologias modernas de produção**. Lavras: Editora UFLA, 181p. 2009.

TRANI, P. E. **Cultura do alho (*Allium sativum*): Diagnóstico e recomendações para seu cultivo no Estado de São Paulo**. 2009. Artigo em Hypertexto. Disponível em: <http://www.infobibos.com/Artigos/2009_2/alho/index.htm>. Acesso em: 13/5/2013.

VIDIGAL, M. C. G.; VIDIGAL FILHO, P. S.; AMARAL JUNIOR, A. T. do; LUCCA E BRACCINI, A. de. Divergência genética entre cultivares de mandioca por meio de estatística multivariada. **Bragantia**, Campinas v. 56, n. 2, p. 263-271, 1997.

VON BERTALANFFY, L. Quantitative laws for metabolism and growth. **Q. Rev. Biol.** 32: 217-231, 1957.

APÊNDICE

A – Dendrogramas das análises de agrupamento

Aqui são apresentados os Dendrogramas obtidos com o algoritmo UPGMA com base na distância generalizada de Mahalanobis para agrupamento dos modelos de regressão não linear Brody (A), Gompertz (B), Logístico (C), Mitscherlich (D) e von Bertalanffy (E), para cada acesso de alho. O número de grupos nas análises foi determinado através do método de Mojena, (1977) com $k = 1,25$. Os dendrogramas da esquerda foram obtidos com um agrupamento pelos avaliadores de qualidade de ajuste, enquanto o da direita utilizam as estimativas dos parâmetros $\hat{\beta}_1$ e $\hat{\beta}_3$.

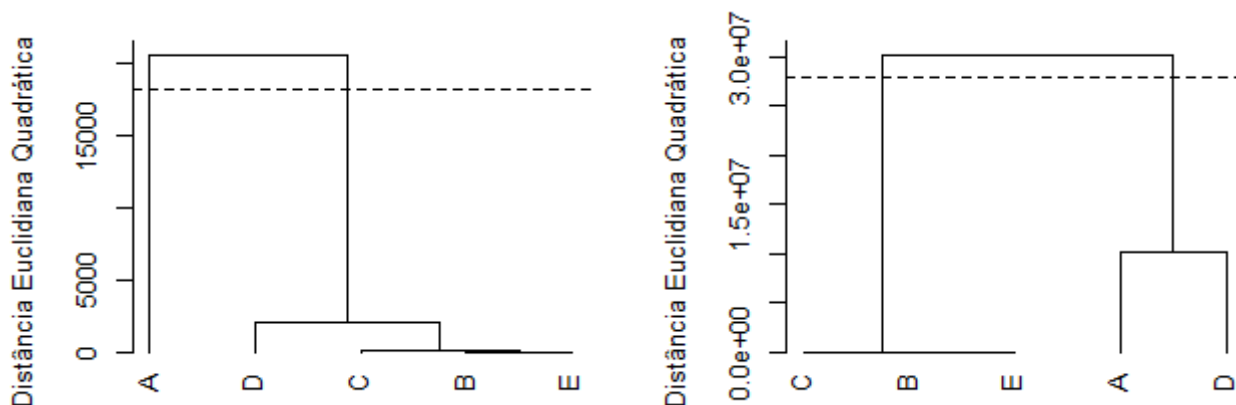


Figura 1A. Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 1

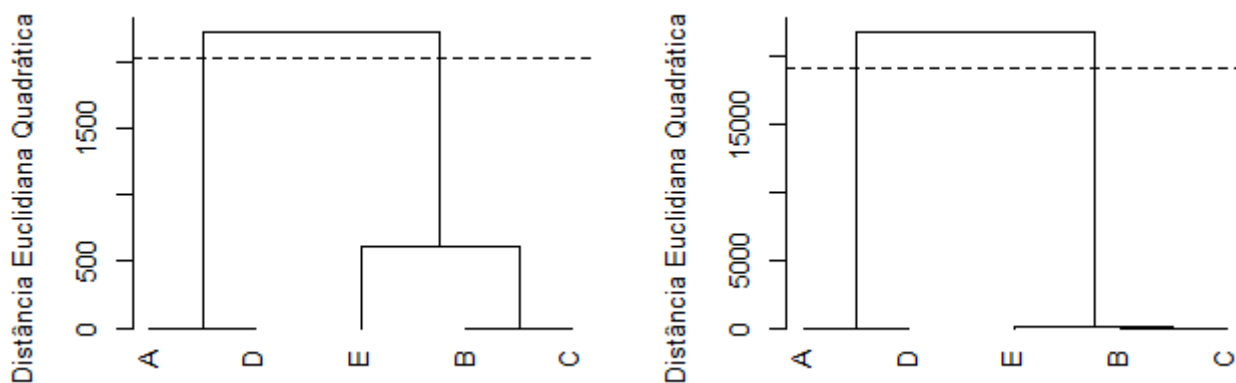


Figura 2A. Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 2

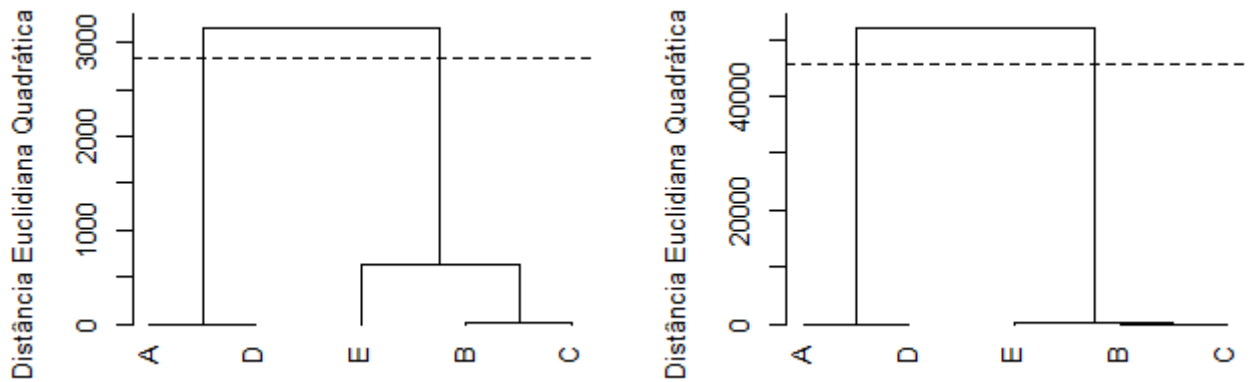


Figura 3A. Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 3.

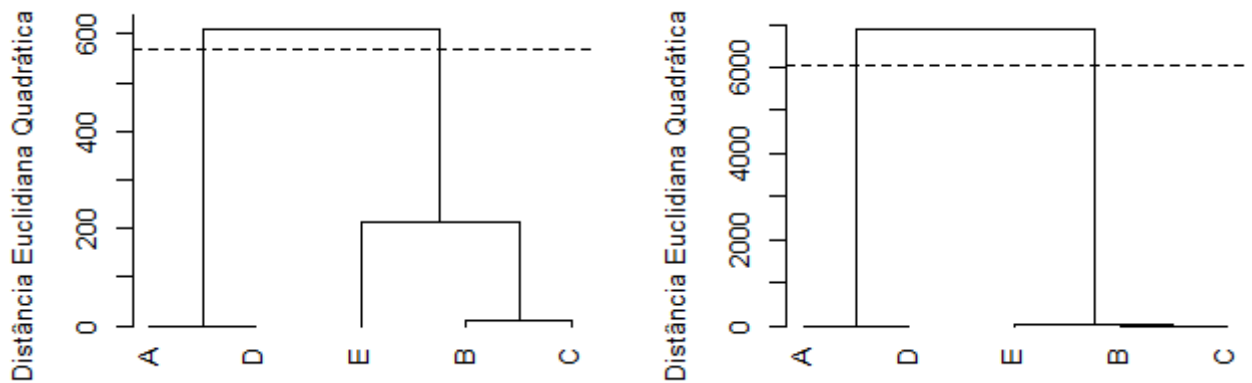


Figura 4A. Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 4

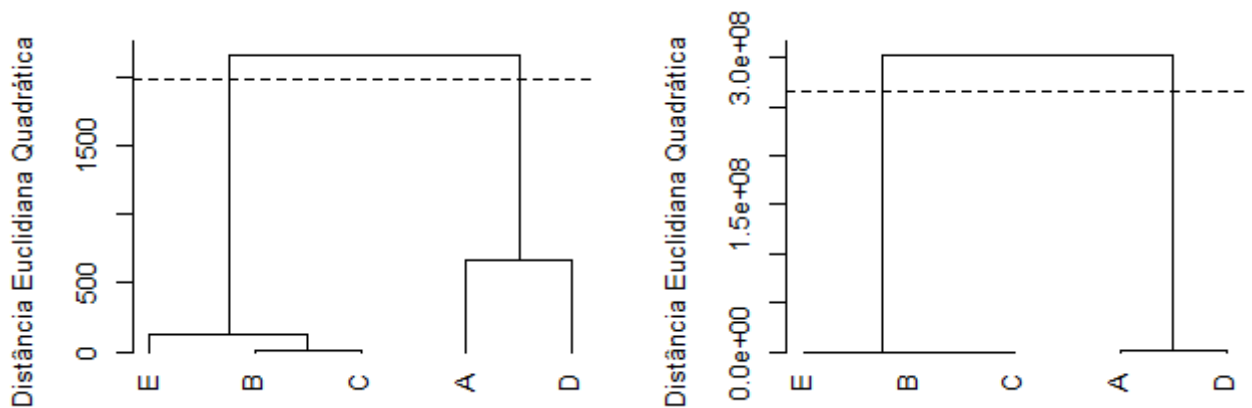


Figura 5A. Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 5

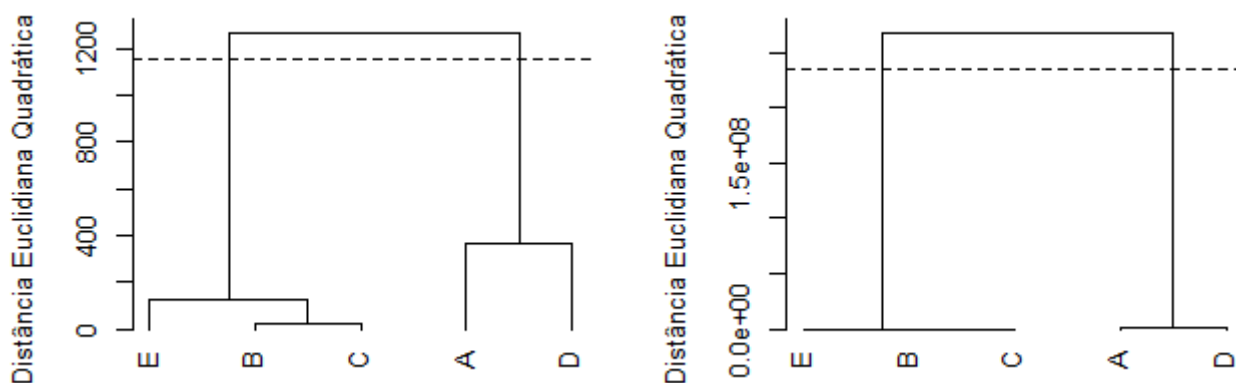


Figura 6A. Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 6

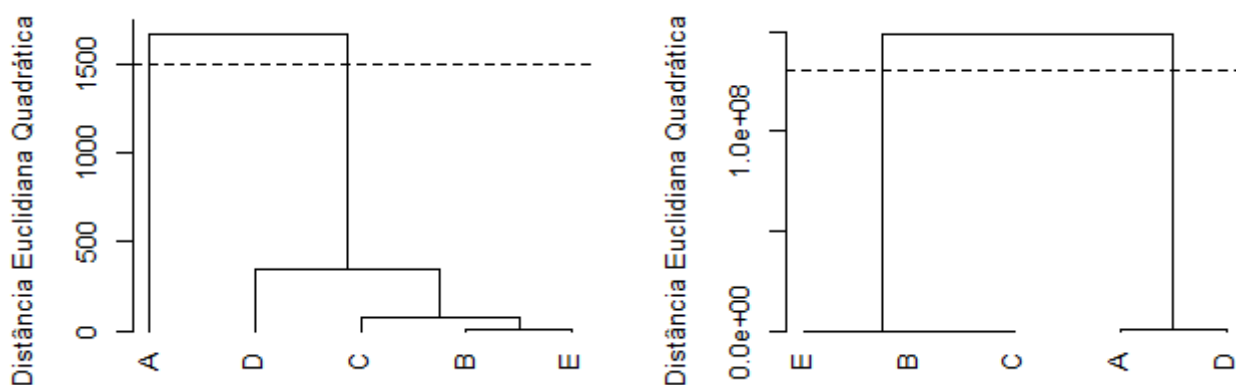


Figura 7A. Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 7

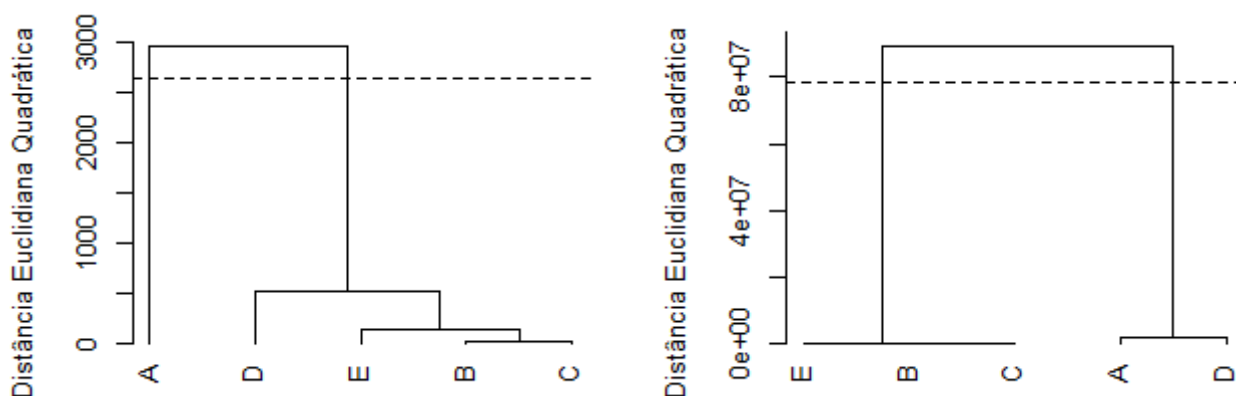


Figura 8A. Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 8

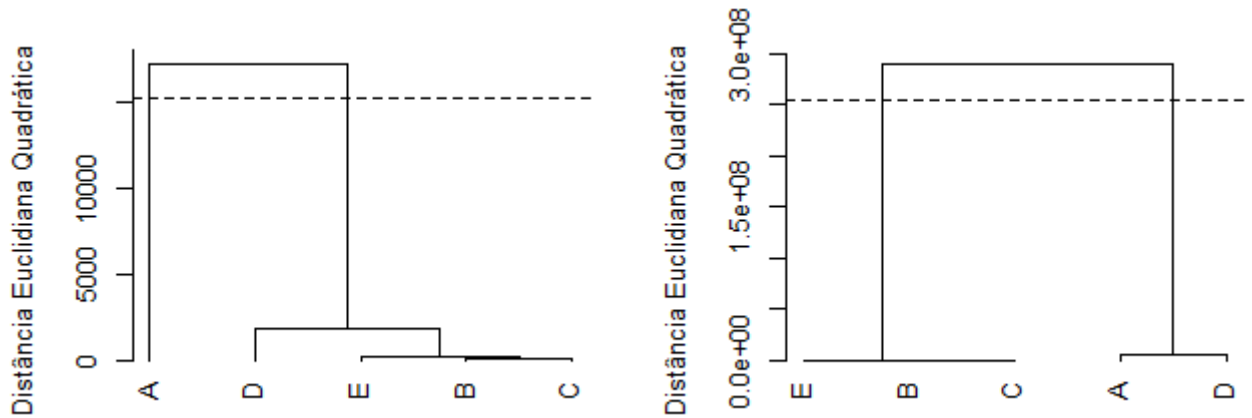


Figura 9A. Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 9

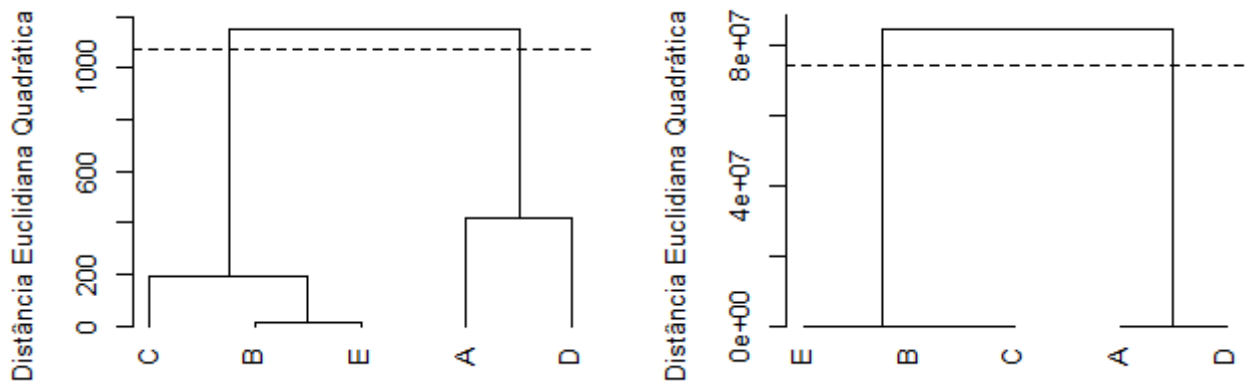


Figura 10A. Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 10

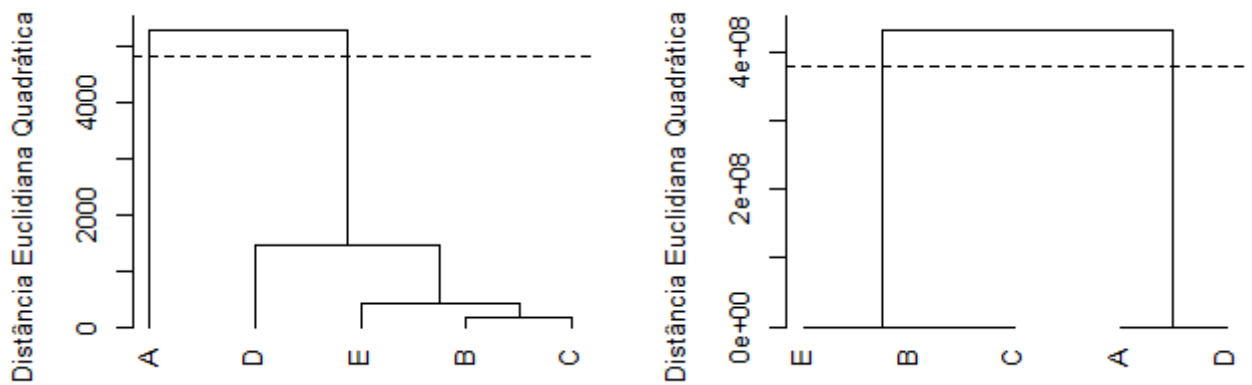


Figura 11A.Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 11

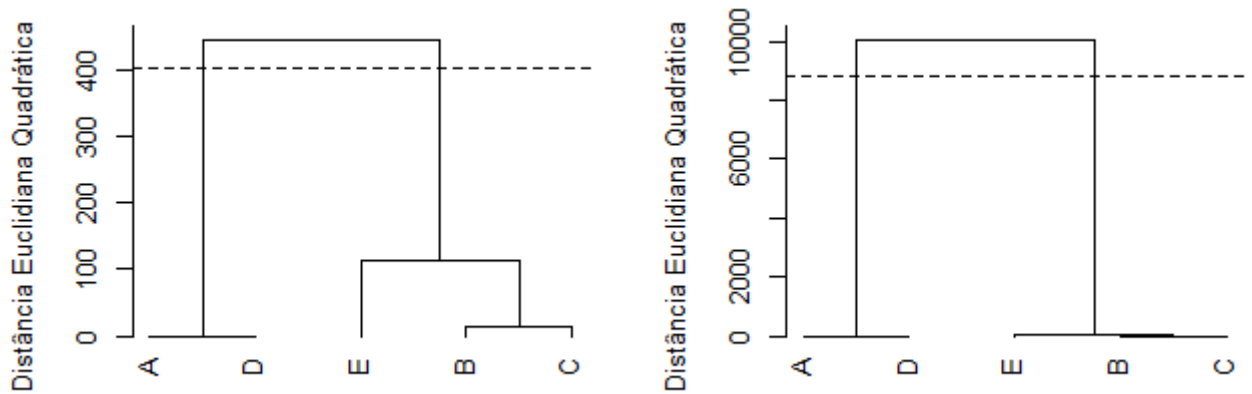


Figura 12A.Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 12

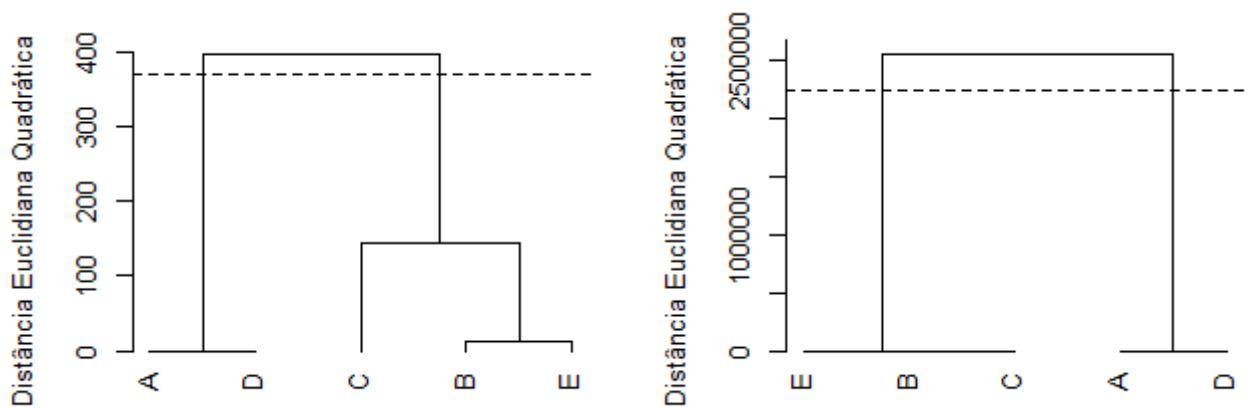


Figura 13A.Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 13

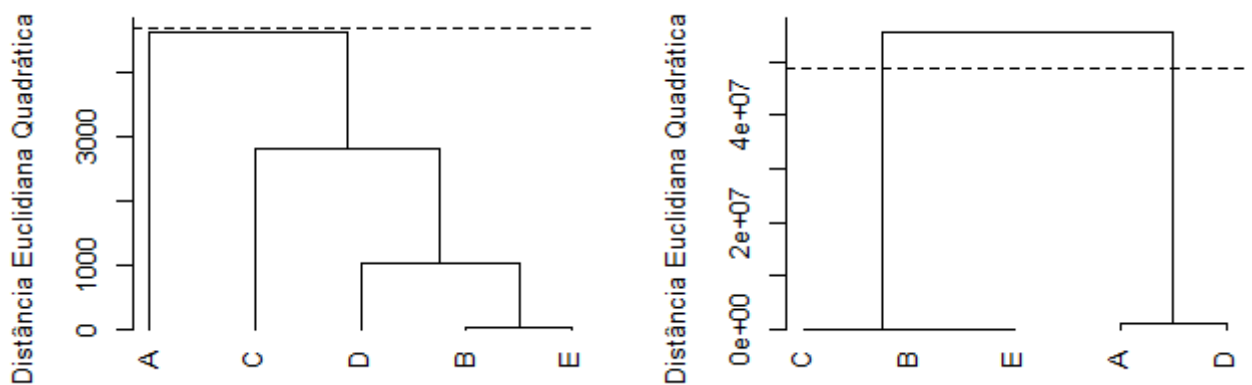


Figura 14A.Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 14

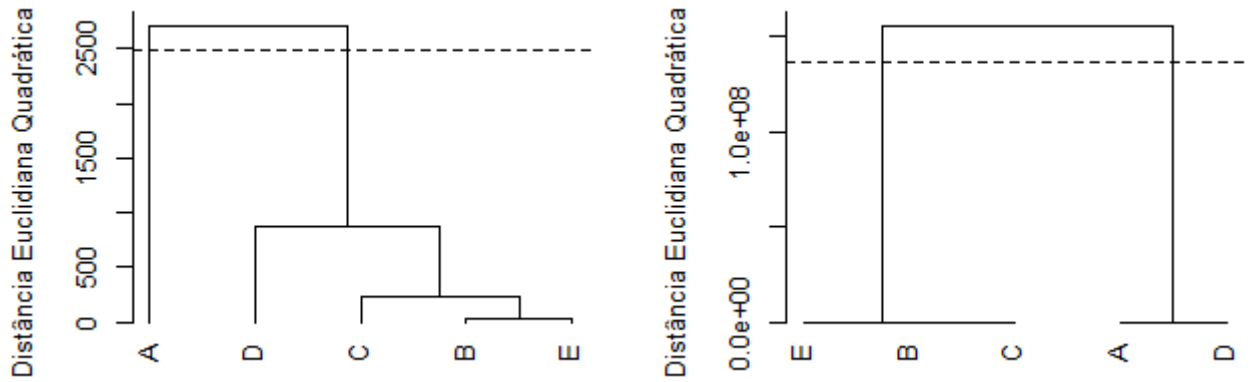


Figura 15A.Dendrogramas obtidos no agrupamento dos cinco modelos de regressão não linear ajustados para o acesso de alho 15

B – Script das análises no *software* R

```
### -----  
### Ajustes dos modelos de regressao nao linear  
### -----  
  
### Entrada de variaveis  
  
# Variavel independente - Dia apos plantio (DAP):  
x<-c(60,90,120,150)  
  
# Variavel dependente - MSTP - medias da materia seca total da  
# planta no acesso N (em gramas):  
y<-c(1.0646, 4.0500, 15.6375, 24.0750)  
# Acesso 1 foi o utilizado no exemplo  
  
### Ajuste dos modelos  
  
# Modelo A - Brody  
mA<-nls(y~A*(1-B*exp(-C*x)), start = list(A=15000,B=51,C=0.05),  
control=nls.control(maxiter=100000,minFactor=1/8192))  
mA  
  
# Modelo B - Gompertz  
mB<-nls(y~A*exp(-B*exp(-C*x)), start = list(A=20,B=450,C=0.06))  
mB  
  
# Modelo C - Logistico  
mC<-nls(y~A/(1+B*exp(-C*x)), start = list(A=22,B=4000,C=0.07))  
mC  
  
# Modelo D - Mitscherlich
```

```

mD<-nls(y~A*(1-exp(B*C-C*x)),start = list(A=6000,B=60,C=0.02),
control=nls.control(maxiter=100000,minFactor=1/8192))
mD

# Modelo E - von Bertalanffy
mE<-nls(y ~ A*(1-B*exp(-C*x))^3,start = list(A=20,B=51,C=0.05))
mE

### -----
### Avaliadores da qualidade do ajuste
### -----

# Utilizando modelo logístico como exemplo
m<-mC

# Coeficiente de determinacao (R2):
SQR <- summary(m)$sigma^2*summary(m)$df[2]
SQT <- var(y)*(length(y)-1)
R2 <- 1-SQR/SQT
R2

# Quadrado medio do residuo (QMR):
QMR <- sum((y- predict(m,x))^2)/(summary(m)$df[2])
QMR

# Desvio medio absoluto do residuo (DMA):
DMA <- sum(abs(y- predict(m,x))/(length(y)))
DMA

# Criterio de informacao de Akaike (AIC):
AIC(m)

# Criterio de informacao Bayesiano (BIC):
BIC(m)

```

```

# Parametros do modelo:
coef(m)

### -----
### Agrupamento metodo UPGMA
### -----

# Utilizando os dados dos avaliadores de qualidade de ajuste do
# acesso 1
dados <- read.table("Acessolaval.txt",header=TRUE)

D <- dist(dados, method = "euclidean")^2
agr <- hclust(D, method = "average")
plot(agr, hang=-1, labels=dados[,1], xlab="", ylab="Distância
Euclidiana Quadrática", main="")
agr$height
mojena <- mean(agr $height)+1.25*sd(agr$height)
k <- length(agr$height[agr$height>mojena]) + 1
abline(h=mojena, v=NULL, col=1, lty=2)
grupos <- cutree(agr, k = k)
grupos

```