

RENATA MACIEL DOS REIS

**MODELOS DE REGRESSÃO NÃO LINEAR PARA DESCRIÇÃO DO
CRESCIMENTO DE 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
2012

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

R375m
2012

Reis, Renata Maciel dos, 1987-

Modelos de regressão não linear para descrição do
crescimento de plantas de alho / Renata Maciel dos Reis.
- Viçosa, MG, 2012.
xiv, 55f. : il. ; 29cm.

Inclui apêndice.

Orientador: Paulo Roberto Cecon.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 41-47

1. Alho - Crescimento - Modelos estatísticos. 2. Análise por
agrupamento. 3. *Allium sativum*. I. Universidade Federal de
Viçosa. II. Título.

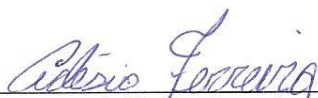
CDD 22. ed. 635.26

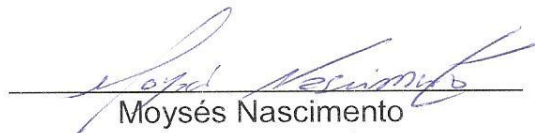
RENATA MACIEL DOS REIS

**MODELOS DE REGRESSÃO NÃO LINEAR PARA DESCRIÇÃO DO
CRESCIMENTO DE 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: 16 de julho de 2012


Adésio Ferreira


Moysés Nascimento
(Coorientador)


Paulo Roberto Cecon
(Orientador)

Aos meus pais, José Sérgio e Terezinha,

Ao meu irmão, Rômulo,

Aos demais familiares,

Aos amigos,

Por terem acreditado em mim,

dedico.

AGRADECIMENTOS

Agradeço primeiramente a Deus, a quem pertence tudo que sou e tenho.

À Universidade Federal de Viçosa, por ter me acolhido como estudante desde minha graduação.

À CAPES e Programa REUNI, pela concessão da bolsa de estudo.

À FAPEMIG, pelo financiamento do projeto de pesquisa.

Aos meus pais e ao meu irmão, pela educação dada e por sempre me apoiarem.

Ao professor e orientador Paulo Roberto Cecon, pela orientação e paciência, pelos inúmeros conhecimentos transmitidos, pelas palavras amigas, pelo exemplo de pessoa e de profissional que é.

Aos professores e coorientadores Moysés Nascimento, Fabyano Fonseca e Silva e Antônio Policarpo Souza Carneiro, pelos ensinamentos e pela atenção durante o mestrado e para a realização deste trabalho.

Aos demais professores do Programa de Pós Graduação em Estatística Aplicada e Biometria, pelos conhecimentos transmitidos durante todo o curso.

Aos professores Mário Puiatti e Fernando Luiz Finger, do Departamento de Fitotecnia da UFV, pela concessão dos dados do experimento.

Ao Anderson, pela paciência e disposição em ajudar sempre que precisei.

Aos colegas do mestrado, pelas horas de estudo e descontração.

Aos amigos, pela confiança em mim depositada e pelo apoio.

A todos que, de alguma forma, contribuíram para que esse trabalho fosse concretizado.

BIOGRAFIA

RENATA MACIEL DOS REIS, filha de José Sérgio dos Reis e Terezinha de Fátima Maciel dos Reis, nasceu em Jesuânia, Minas Gerais, no dia 22 de junho de 1987.

Em maio de 2006, ingressou no curso de Matemática da Universidade Federal de Viçosa, Viçosa – MG, graduando-se em julho de 2010.

Em agosto de 2010 iniciou o curso de Mestrado no Programa de Pós Graduação em Estatística Aplicada e Biometria na Universidade Federal de Viçosa, submetendo-se à defesa de dissertação em 16 de julho de 2012.

SUMÁRIO

LISTA DE FIGURAS	viii
LISTA DE TABELAS	ix
RESUMO	xi
ABSTRACT	xiii
1. INTRODUÇÃO	1
2. REVISÃO BIBLIOGRÁFICA.....	4
2.1. A Cultura do alho	4
2.2. Divergência genética	5
2.2.1. Medidas de dissimilaridade	6
2.2.2. Análise de agrupamento	7
2.2.3. Métodos de agrupamento	7
2.3. Modelos de regressão não linear	8
2.3.1. Avaliadores da qualidade do ajuste	10
3. MATERIAL E MÉTODOS.....	12
3.1. Descrição do experimento	12
3.2. Análise de agrupamento para os acessos.....	14
3.2.1. Dissimilaridade entre acessos.....	14
3.2.2. Método da variância mínima de Ward	15
3.2.3. Determinação do número de grupos	18

3.3. Modelos de regressão não linear	19
3.3.1. Método dos mínimos quadrados ordinários	21
3.3.2. Processo iterativo de Gauss-Newton	23
3.3.3. Avaliadores da qualidade do ajuste.....	26
3.3.3.1. Coeficiente de determinação (R^2).....	27
3.3.3.2. Quadrado médio do resíduo (QMR).....	27
3.3.3.3. Desvio médio absoluto dos resíduos (DMA).....	28
4. RESULTADOS E DISCUSSÃO	29
4.1. Agrupamento dos acessos	29
4.2. Estimativas dos parâmetros dos modelos de regressão não linear e avaliadores da qualidade do ajuste	31
5. CONCLUSÕES.....	40
6. REFERÊNCIAS	41
APÊNDICE.....	48
A – Script das análises de agrupamento no <i>Software R</i>	49
B – Script dos ajustes dos modelos no <i>Software R</i>	51

LISTA DE FIGURAS

- Figura 1.** Dendrograma obtido com o algoritmo de Ward com base na distância generalizada de Mahalanobis (D^2) para agrupamento de 20 acessos de alho pertencentes ao BGH/UFV. UFV, Viçosa (MG), 201229
- Figura 2.** Curvas de acúmulo de matéria seca do bulbo (MSB) ajustadas pelo modelo Logístico dos três grupos de acessos de alho. UFV, Viçosa (MG), 201233
- Figura 3.** Curvas de acúmulo de matéria seca da raiz (MSR) ajustadas pelo modelo Logístico dos três grupos de acessos de alho. UFV, Viçosa (MG), 201236
- Figura 4.** Curvas de acúmulo de matéria seca total da planta (MSTP) ajustadas pelo modelo Logístico dos três grupos de acessos de alho. UFV, Viçosa (MG), 201238

LISTA DE TABELAS

- Tabela 1.** Relação dos 20 acessos de alho registrados no Banco de Germoplasmas de Hortaliças da Universidade Federal de Viçosa (BGH/UFV). UFV, Viçosa (MG), 2012.....13
- Tabela 2.** Modelos de regressão não linear para descrever acúmulo de matéria seca20
- Tabela 3.** Formação dos grupos de 20 acessos de alho do BGH/UFV obtidos com o algoritmo de Ward com base na distância generalizada de Mahalanobis (D^2). UFV, Viçosa (MG), 201230
- Tabela 4.** Valores médios das características de matéria seca do bulbo (MSB), da raiz (MSR) e total da planta (MSTP), em gramas, dos respectivos grupos e dias após plantio. UFV, Viçosa (MG), 2012.....30
- Tabela 5.** Estimativas dos parâmetros ($\hat{\beta}_1$, $\hat{\beta}_2$ e $\hat{\beta}_3$), coeficiente de determinação (R^2), quadrado médio do resíduo (QMR) e desvio médio absoluto dos resíduos (DMA), de acordo com os modelos estudados para as médias de matéria seca do bulbo (MSB) de cada grupo de acesso. UFV, Viçosa (MG), 2012.....32

Tabela 6. Estimativas dos parâmetros ($\hat{\beta}_1$, $\hat{\beta}_2$ e $\hat{\beta}_3$), coeficiente de determinação (R^2), quadrado médio do resíduo (QMR) e desvio médio absoluto dos resíduos (DMA), de acordo com os modelos estudados para as médias de matéria seca da raiz (MSR) de cada grupo de acesso. UFV, Viçosa (MG), 2012.....35

Tabela 7. Estimativas dos parâmetros ($\hat{\beta}_1$, $\hat{\beta}_2$ e $\hat{\beta}_3$), coeficiente de determinação (R^2), quadrado médio do resíduo (QMR) e desvio médio absoluto dos resíduos (DMA), de acordo com os modelos estudados para as médias de matéria seca total da planta (MSTP) de cada grupo de acesso. UFV, Viçosa (MG), 201237

RESUMO

REIS, Renata Maciel dos, M.Sc., Universidade Federal de Viçosa, julho de 2012. **Modelos de regressão não linear para descrição do crescimento de plantas de alho.** Orientador: Paulo Roberto Cecon. Coorientadores: Moysés Nascimento, Fabyano Fonseca e Silva e Antônio Policarpo Souza Carneiro.

O objetivo deste estudo foi escolher um modelo de regressão não linear que melhor descreve o acúmulo de matéria seca de diferentes partes da planta do alho ao longo do tempo (60, 90, 120 e 150 dias após plantio). Foram utilizados 20 acessos de alho pertencentes ao Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH/UFV). A fim de se trabalhar apenas com grupos de acessos semelhantes, aplicou-se a análise de agrupamento para a formação desses grupos. As matérias secas da folha, do pseudocaule, do bulbo e da raiz foram definidas como as variáveis nessa análise de agrupamento, que foi realizado por meio do algoritmo de Ward, utilizando como medida de dissimilaridade a distância generalizada de Mahalanobis. O número ótimo de grupos foi determinado por meio do Método de Mojena, o qual indicou três grupos de acessos, cujas médias de matéria seca do bulbo, da raiz e total da planta foram utilizadas para o ajuste de sete modelos de regressão não linear, a saber: Mitscherlich, Gompertz, Logístico, Meloun I, Meloun II, Brody e von Bertalanffy. A identificação do modelo que melhor se ajustou as três características de cada grupo foi realizada mediante coeficiente de determinação (R^2), o quadrado médio do

resíduo (QMR) e o desvio médio absoluto dos resíduos (DMA). Comparando os valores desses avaliadores observou-se que, para as três características dos três grupos, o modelo que melhor se ajustou foi o modelo Logístico.

ABSTRACT

REIS, Renata Maciel dos, M.Sc., Universidade Federal de Viçosa, July, 2012. **Nonlinear regression models for the growth description of plants garlic**. Adviser: Paulo Roberto Cecon. Co-advisers: Moysés Nascimento, Fabyano Fonseca e Silva and Antônio Policarpo Souza Carneiro.

The objective of this study was to choose a nonlinear regression model that best described the dry matter accumulation in different parts of the plant garlic over time (60, 90, 120 and 150 days after planting). Were used 20 garlic accessions belonging to the Vegetable Germplasm Bank of Universidade Federal de Viçosa. In order to work only with groups of similar accessions, was applied the cluster analysis in order to obtaining these clusters. The dry matter of leaf, pseudostem, bulb and root were defined as variables in this cluster analysis, which was conducted by the Ward algorithm, using as dissimilarity measure the Mahalanobis distance. Based on Mojena's method to determine the number of groups, was formed three groups of accessions, whose means of dry matter of bulb, of root and of the whole plant were used to fitting of seven nonlinear regression models, namely : Mitscherlich, Gompertz, Logistic, Meloun I Meloun II, von Bertalanffy and Brody. Aiming to choose the best fitted model to the three characteristics of each group were calculated coefficient of determination (R^2), the error mean square (EMS) and the average deviation absolut error. Comparing the

values of the evaluators found that, for the three characteristics of the three groups, the best fit model was the Logistic model.

1. INTRODUÇÃO

O alho (*Allium sativum* L.) é uma hortaliça que se destaca na culinária mundial e, particularmente, na brasileira, por possuir características acentuadas de aroma e sabor que lhe atribuem propriedades condimentares e também por suas propriedades medicinais (MENEZES SOBRINHO et al., 1999; SOUZA et al., 2007). Apresenta forte importância econômica e, além disso, tem grande influência no sistema de agricultura familiar, por empregar mão-de-obra desde o plantio até a colheita (TRANI, 2009). Em relação a sua diversidade genética, existem inúmeros cultivares de alho no país que surgiram, segundo Souza et al. (1977), principalmente por mutações somáticas e seleções de características desejáveis pelos agricultores. Dessa forma evidencia-se a necessidade de se conduzir pesquisas com o objetivo de estudar esta divergência genética do alho.

Na literatura são apresentadas várias técnicas para estudo de divergência genética, tais como componentes principais, variáveis canônicas e de análises de agrupamento, como feito por Menezes Sobrinho (1999) e Silva (2012) para acessos (variedades) de alho e por Cargnelutti Filho et al. (2010) e Carvalho et al. (2003) para outras culturas como o feijão e o algodão, respectivamente. Para que se avalie a divergência genética de forma simultânea em relação às diversas características de determinada cultura é recomendado que se utilize medidas de dissimilaridade (CRUZ e CARNEIRO, 2003), para serem utilizadas em análises de agrupamento, que tem por finalidade reunir objetos em grupos, de forma que exista máxima

homogeneidade dentro do grupo e máxima heterogeneidade entre os grupos (JOHNSON e WICHERN, 1992).

E mais, para que se faça um manejo adequado por meio de informações relevantes e também para que se detecte possíveis problemas no desenvolvimento das culturas recomenda-se estudar a trajetória de crescimento de plantas e o acúmulo de matéria seca de suas diferentes partes. E, segundo Silva et al. (2001), o interesse nesse estudo vem aumentando nesses últimos anos juntamente com o desenvolvimento de novas técnicas computacionais que fornecem maior agilidade e precisão das análises.

Para realizar esse estudo, tem-se utilizado modelos de regressão linear e não linear para descrever, tanto o crescimento da planta, quanto o acúmulo de matéria seca ao longo do tempo (UEDA et al., 2010; PÔRTO et al., 2007; ALVAREZ e CASTRO, 1999). Os modelos de regressão não linear proporcionam, segundo Maia et al. (2009), vantagens em relação aos de regressão linear, pois estes podem auxiliar o pesquisador no processo de tomada de decisão já que apresentam parâmetros com interpretação biológica, como visto nos trabalhos de Oliveira et al. (2000) e Santoro et al. (2005).

Diante dos variados modelos de regressão não linear encontrados na literatura para descrever curvas de crescimento e acúmulos de matéria seca de plantas, a tarefa de definir o modelo que melhor se ajuste é relatada como árdua (OLIVEIRA et al., 2000; SANTORO et al., 2005, SILVEIRA et al., 2011). Dessa forma, são propostos alguns critérios para se avaliar um modelo, seja por meio de análise gráfica ou por estatísticas citadas na literatura (SILVEIRA et al., 2011; SARMENTO et al., 2006; OLIVEIRA et al., 2000).

Em face destes contextos e visto que esse tipo de estudo com a cultura do alho e suas variedades é pouco difundido, este trabalho teve por objetivo escolher um modelo de regressão não linear que melhor descreve o acúmulo de matéria seca de diferentes partes da planta do alho ao longo do tempo, como a matéria seca do bulbo, da raiz e matéria seca total da planta. Os dados foram provenientes de 20 acessos de alho pertencentes ao Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa

(BGH/UFV) e, a fim de se trabalhar apenas com grupos de acessos semelhantes, aplicou-se análise de agrupamento para a formação desses grupos.

2. REVISÃO BIBLIOGRÁFICA

2.1. A cultura do alho

Segundo Trani (2009), o alho é uma planta herbácea de propagação vegetativa, pertencente à família *Alliaceae*. Apresenta folhas lanceoladas que diferem dependendo da localização na planta e possui pseudocaule formado pelas bainhas das folhas. Em condições climáticas favoráveis as gemas do caule se desenvolvem formando cada uma um bulbilho (dente), que em conjunto forma o bulbo (cabeça).

O Brasil, tradicional importador de alho, vem reduzindo sua dependência externa graças ao desenvolvimento de novas tecnologias e expansão do cultivo de alhos nobres (FEITOSA et al., 2009). O alho é uma das mais importantes hortaliças cultivadas no Brasil e, de acordo com o IBGE (2012), na safra de 2011 o país produziu 142.494 toneladas em 12.836 hectares, sendo Goiás o maior produtor (46.700 toneladas) com área plantada de 3.096 hectares e Minas Gerais o segundo maior produtor (40.960 toneladas) com área plantada de 3.075 hectares. O rendimento médio foi estimado em 11.101 kg/ha.

Com esse aumento no interesse pela cultura do alho no país, aumenta também a necessidade de estudos que visem indicar um manejo adequado por meio de informações relevantes e detectar possíveis problemas no desenvolvimento das culturas, para isso é recomendado estudar a trajetória de crescimento de plantas e de características

relacionadas a estas, tais como matéria seca de suas diferentes partes. Uma forma de se fazer esse estudo é utilizando modelos de regressão não linear para descrever, tanto o crescimento da planta, quanto o acúmulo de matéria seca ao longo do tempo, visto que esses tipos de modelos proporcionam vantagens em relação aos de regressão linear, pois podem auxiliar o pesquisador no processo de tomada de decisão já que apresentam parâmetros com interpretação biológica. (MAIA et al., 2009). E mais, a utilização de curvas de acúmulo de matéria seca e de nutrientes para diversas cultivares de hortaliças, como um parâmetro para a recomendação de adubação, é uma boa indicação da demanda em cada etapa do desenvolvimento da planta (PÔRTO et al., 2007).

A cultura é exigente em baixa temperatura e fotoperíodo longo para a bulbificação. Em geral, temperaturas médias de 12,8 a 23,9 °C favorecem o desenvolvimento normal do alho. Porém, existem diferenças entre os cultivares quanto à resposta à temperatura e ao fotoperíodo com consequente variação de resposta destas às regiões de cultivo e épocas de plantio (MUELLER et al., 1990). Em consequência dessas diferenças e de sua divergência genética, num conjunto de variados acessos, pode ser que hajam acessos mais semelhantes e outros mais divergentes. Assim, para que se possa trabalhar apenas com grupos de acessos semelhantes, uma alternativa é utilizar técnicas de análise de agrupamento.

2.2. Divergência genética

A divergência genética entre e dentro de populações encontradas em suas condições naturais, em bancos de germoplasma ou desenvolvidas nos programas de melhoramento genético pode ser predita pelas diferenças entre os valores fenotípicos mensurados em suas unidades (indivíduos, famílias, etc.) (DINIZ FILHO, 2000 *apud* CRUZ et al., 2011).

Pode-se inferir sobre a diversidade genética de duas formas, a primeira é de natureza quantitativa e a outra de natureza preditiva (CRUZ e CARNEIRO, 2003). Os métodos preditivos de diversidade genética tem sido bastante utilizados, sobretudo pelo fato de que, ao se basearem 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.

Estudos de divergências genéticas tem sido realizados em diferentes culturas como pimenta (FARIA, 2009), algodão (CARVALHO et al., 2003), café (FONSECA et al., 2006) e no alho (SILVA, 2012).

2.2.1. Medidas de dissimilaridade

Medida de dissimilaridade pode ser entendida como um critério que quantifique a distância entre dois objetos, assim, quanto menores os valores destas medidas, mais similares são os objetos comparados. Um grande número de distâncias tem sido proposto e usado em análise de agrupamento e, segundo Mingoti (2005), cada uma delas produz um determinado tipo de agrupamento.

O termo dissimilaridade surgiu devido à relação da distância entre dois objetos, a medida que essa distância cresce, a divergência entre eles aumenta, ou seja, tornam-se cada vez mais dissimilares.

Segundo Cruz e Carneiro (2003), as seguintes medidas de dissimilaridade são comumente utilizadas: Distância Euclidiana, Distância Euclidiana Média, Quadrado da Distância Euclidiana Média e Distância generalizada de Mahalanobis. Johnson e Wichern (1992) comentam que a distância Euclidiana é insatisfatória para algumas situações estatísticas, pois a contribuição de cada coordenada tem o mesmo peso para o cálculo da distância. Assim, se aconselha utilizar a distância euclidiana somente quando as características avaliadas não tiverem graus de correlação residual significativos. Já a distância generalizada de Mahalanobis (1936) leva em consideração as variâncias e covariâncias residuais, podendo ser utilizada em casos que se verifique correlação residual.

2.2.2. Análise de agrupamento

A análise de agrupamento é uma maneira de se obter grupos homogêneos, por um esquema que possibilite reunir os indivíduos em um determinado número de grupos, de modo que exista grande homogeneidade dentro de cada grupo e heterogeneidade entre eles (JOHNSON e WICHERN, 1992). Segundo Manly (1994), uma análise de agrupamento utiliza os valores das variáveis para esquematizar formas de agrupar os objetos em classes, de modo que objetos similares estejam na mesma classe. O autor afirma também que uma análise de agrupamento pode ser realizada tanto para encontrar os “verdadeiros” grupos dos objetos quanto visando reduzir os dados.

Essa metodologia visa dividir um grupo original de observações em vários grupos, de acordo com algum critério de similaridade ou de dissimilaridade (CRUZ et al., 2004). Tais critérios constituem a base desta metodologia, visto que a análise de agrupamento se inicia com a obtenção de uma matriz de medidas de similaridade ou dissimilaridade obtidas da matriz de dados.

2.2.3. Métodos de agrupamento

Segundo Cruz e Carneiro (2003), há inúmeros métodos de agrupamento que se distinguem pelo tipo de resultado a ser fornecido e pelas diferentes formas de se definir a proximidade entre um indivíduo e um grupo já formado ou entre dois grupos quaisquer. Em todos os casos não se conhece, *a priori*, o número de grupos a serem estabelecidos, e diferentes métodos proporcionam diferentes resultados.

Os tipos mais utilizados de agrupamento são os de otimização e os hierárquicos. O objetivo dos métodos de otimização é alcançar uma partição dos indivíduos que otimize (maximize ou minimize) alguma medida predefinida.

Os métodos hierárquicos agrupam os objetos por um processo que se repete em vários níveis, até que seja estabelecido um diagrama

bidimensional em forma de árvore, que ilustra as fusões ou partições efetuadas em cada nível sucessivo do processo de agrupamento denominado dendrograma (Cruz e Carneiro, 2003). Os métodos hierárquicos mais conhecidos são: método do vizinho mais próximo, método do vizinho mais distante, método UPGMA, o método do centróide e o método da variância mínima de Ward.

Segundo Rosemburg (1984), o método de Ward tem sido o método hierárquico preferido em alguns casos, devido a visualização de grupos bem definidos no dendrograma.

O número de grupos nos métodos hierárquicos pode ser definido por alguns critérios: razões práticas do pesquisador; análise visual das ramificações do dendrograma, sendo o corte estabelecido em pontos onde há mudança abrupta da ramificação; e critérios estatísticos, como por exemplo: desvio padrão médio e coeficiente de determinação, propostos por Khattree e Naik (2000, apud Silva, 2012); método de Mojena (1977), baseado no tamanho relativo dos níveis de fusões (distâncias) no dendrograma.

2.3. Modelos de regressão não linear

A palavra regressão foi incorporada na linguagem estatística por Pearson para designar a expressão de uma variável como função de uma ou mais variáveis (CECON et al., 2012). Segundo Mazucheli e Achcar (2002), avaliar a possível relação entre uma variável dependente com uma ou mais variáveis independentes é uma das tarefas mais comuns em análise estatística. Pode-se atingir este objetivo por meio dos bem conhecidos modelos de regressão, os quais se dividem em duas classes distintas: modelos de regressão linear e modelos de regressão não linear. A primeira classe é assim chamada quando sua variável resposta (dependente) é expressa como função linear dos coeficientes de regressão, já os modelos de regressão não linear são aqueles em que pelo menos uma das derivadas parciais em relação aos parâmetros está em função de pelo menos um parâmetro. (RATKOWSKY, 1990).

Modelos de regressão linear tem aplicações nas mais diversas áreas do conhecimento. Entretanto, existem muitas situações em que esses tipos de modelos podem não ser apropriados. Em aplicações mais realistas, especialmente nos casos de crescimento biológico, pode ser necessário ajustar funções não lineares para melhor explicar o processo de crescimento. Ainda, segundo o autor, aplicações de modelos de regressão não linear são típicas em Biologia, quando plantas, ou parte delas, e organismos crescem com o tempo ou a idade, mas há também aplicações em Economia e Engenharia para esses tipos de modelos (REGAZZI, 2003).

Um modelo de regressão não linear é considerado “intrinsecamente linear” se este pode ser reduzido a um modelo linear por meio de uma reparametrização apropriada. Pode-se, ainda, usar o termo “intrinsecamente linear” para referir-se a modelos que podem ser linearizados via alguma transformação. Caso não seja possível obter uma reparametrização ou uma transformação apropriada, que reduza o modelo à forma linear, tem-se os chamados modelos “intrinsecamente não lineares”. (MAZUCHELI e ACHCAR, 2002).

De acordo com Souza (1998), modelos de regressão não linear podem ser escrito da seguinte forma:

$$y_i = f(x_i, \boldsymbol{\theta}) + \varepsilon_i \quad i = 1, \dots, n \quad (1)$$

onde a função resposta $f(x, \boldsymbol{\theta})$ é não linear, y_i representa a observação da variável dependente, x_i representa um vetor de observações em k variáveis preditoras, $\boldsymbol{\theta} = [\theta_1 \ \dots \ \theta_p]'$ é um vetor de parâmetros p dimensional desconhecido e ε_i representa o erro aleatório.

Encontrar o estimador do parâmetro que minimize a soma dos quadrados residuais é uma forma de ajustar o modelo, esse método é chamado de Método dos Mínimos Quadrados Ordinários. Segundo Ratkowsky (1990), o método dos mínimos quadrados ordinários para regressão não linear, ao contrário do método para regressão linear, não apresenta expressões matemáticas explícitas. Dessa forma, algum método iterativo deve ser utilizado para obter as estimativas de mínimos quadrados.

Diversos métodos iterativos são propostos na literatura, os mais utilizados, segundo Bates e Watts (1988), são o método de Gauss-Newton ou método da linearização, o método *Steepest-Descent* ou método do gradiente e o método de Marquardt.

Dentre os modelos mais utilizados na literatura para descrever curvas de crescimento de animais e plantas e também acúmulo de matéria seca estão os modelos Logístico, Gompertz, von Bertalanffy, Brody, Mitscherlich, Richard (SANTORO et al., 2005; SARMENTO et al., 2006; PÔRTO et al., 2007; BRITO et al., 2007; MAIA et al., 2009; SILVEIRA, 2010; entre outros).

2.3.1. Avaliadores da qualidade do ajuste

Quando diferentes modelos de regressão não linear são ajustados a um mesmo conjunto de dados, é importante que se utilize de ferramentas estatísticas visando compará-los e indicar o melhor modelo. Essas ferramentas são denominadas avaliadores da qualidade do ajuste e, encontra-se na literatura, diversas propostas para os mesmos (SILVEIRA, 2010).

Na literatura especializada, a maioria dos trabalhos envolvendo comparação modelos de crescimento utiliza apenas o coeficiente de determinação ajustado (R_{aj}^2) e o desvio padrão residual como avaliadores da qualidade do ajuste que, segundo Silveira (2010), tal procedimento pode ser explicado pelo fato destes avaliadores serem apresentados diretamente nos arquivos de saídas de softwares estatísticos como SAS® (SAS, 2003) e R (R DEVELOPMENT CORE TEAM, 2011).

Além destes, outros avaliadores como Quadrado médio do resíduo (QMR), coeficiente de determinação (R^2) porcentagem de convergência ($CG\%$), Critério de Informação de Akaike (AIC), Critério de informação Bayesiano (BIC), Desvio médio absoluto (DMA) podem ser encontrados na literatura, em que Agostinnetto et al. (2004), Sarmento et al. (2006) e Maia et al. (2009) utilizam o QMR e R^2 , Maia et al. (2009) utilizam também o $CG\%$,

Silveira et al. (2011) utilizam o *AIC* e *BIC* e Sarmento et al. (2006) utilizam também o *DMA*.

3. MATERIAL E MÉTODOS

3.1. Descrição do experimento

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

O delineamento experimental foi o de blocos completos casualizados com quatro repetições. Foram avaliados 20 acessos de alho registrados no Banco de Germoplasma de Hortaliças (BGH/UFV).

A avaliação de matéria seca das partes das plantas desses acessos foi realizada em quatro períodos:

- i) Primeiro período e inicial (60 dias após plantio, dap);
- ii) Segundo período (90 dap);
- iii) Terceiro período (120 dap);
- iv) Quarto período e final (150 dap).

As seguintes matérias secas foram avaliadas: da folha (MSF), do pseudocaule (MSP), do bulbo (MSB), da raiz (MSR) e matéria seca total da planta (MSTP).

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. A MSTP foi obtida pelo somatório da MSF, MSP, MSB e MSR.

Os 20 acessos utilizados nesse trabalho são apresentados na Tabela 1 com seus respectivos nomes comuns e origem.

Tabela 1. Relação dos 20 acessos de alho registrados no Banco de Germoplasmas de Hortaliças da Universidade Federal de Viçosa (BGH/UFV). UFV, Viçosa (MG), 2012.

Nº	Acesso	Origem	Nome comum
1	4400	Capim Branco – MG	Alho Roxão de Capim Branco
2	4483	Florestal – MG	Barbado do Rio Grande
3	4499	Florestal – MG	Montes Claros
4	4500	Florestal – MG	Sacaia de Guanhões
5	4825	Rio Pantojas – ES	Sapé
6	4827	S ^{ta} Maria - S ^{ta} Leopoldina	-
7	4837	Vargem Alegre – MG	-
8	5366	Brasília – DF	Cultura de tecidos
9	5934	Viçosa – MG	Amarante
10	5936	Viçosa – MG	Centenário
11	5947	Viçosa – MG	Cateto Roxo Seleção
12	5949	Viçosa – MG	Cajuru
13	5950	Viçosa – MG	Chinês – 1
14	5952	Viçosa – MG	Chinês – 4
15	5961	-	Peruano
16	5966	Nova Venécia – ES	Cateto Roxo (d ^t arredondado)
17	6394	-	Amarante Novo Cruzeiro
18	6397	Presidente Médice	Presidente Médice
19	7620	Espirito Santo – ES	Amarante – EPAMIG
20	7622	Espirito Santo – ES	Amarante – EPAMIG

3.2. Análise de agrupamento para os acessos

Utilizou-se a técnica de análise de agrupamento aplicada às médias das quatro repetições dos acessos para avaliar a divergência genética entre eles. Assim, confeccionou-se um conjunto de dados multivariados, no qual os 20 acessos de alho registrados no Banco de Germoplasma de Hortaliças (BGH/UFV) correspondem aos indivíduos e a matéria seca da folha, do pseudocaule, do bulbo e da raiz aos 150 dias após plantio correspondem às variáveis. A matéria seca total da planta é uma combinação linear das demais matérias secas, por esse motivo ela não foi considerada uma variável para a análise de agrupamento.

3.2.1. Dissimilaridade entre acessos

A medida de dissimilaridade adotada foi a distância generalizada de Mahalanobis (D^2), pois se pressupôs que as características utilizadas como variáveis na análise de agrupamento apresentassem certo grau de correlação, não sendo totalmente independentes e, segundo Cruz e Carneiro (2003), a distância generalizada de Mahalanobis apresenta a vantagem em relação as distâncias euclidianas pois leva em consideração a magnitude das variâncias e covariâncias residuais que existem entre as características, possíveis de serem estimadas a partir de informações dos acessos avaliados em delineamentos experimentais.

As distâncias de Mahalanobis são obtidas por meio da expressão:

$$D_{ii'}^2 = \delta_{ii'}' \Psi^{-1} \delta_{ii'}$$

em que:

$D_{ii'}^2$: é a distância generalizada de Mahalanobis entre os acessos i e i' ;

$\delta_{ii'}' = [d_1 \quad d_2 \quad \cdots \quad d_v]$, sendo $d_j = Y_{ij} - Y_{i'j}$ para cada par de acessos (i, i') ;

Y_{ij} : é a média do i -ésimo acesso em relação à j -ésima variável,
 $j = 1, 2, \dots, v$

Ψ^{-1} : é a inversa da matriz de variâncias e covariâncias residuais, cujos elementos da diagonal principal são os quadrados médios do resíduo (QMR) da análise de variância e os elementos fora dela são os produtos médios do resíduo (PMR) entre os pares de variáveis, podendo ser obtido por:

$$PMR(Y_1, Y_2) = \frac{QMR(Y_1 + Y_2) - QMR(Y_1) - QMR(Y_2)}{2}$$

uma vez que se verifica a relação:

$$Cov(Y_1, Y_2) = \frac{V(Y_1 + Y_2) - V(Y_1) - V(Y_2)}{2}$$

Além de se considerar as variâncias e covariâncias residuais das variáveis pela distância generalizada de Mahalanobis, o procedimento de agrupamento foi realizado considerando-se as escalas padronizadas das variáveis, por meio de: Y_{ij}/s_j , em que s_j é o desvio padrão da variável Y_j .

Nesse trabalho a matriz de distâncias de Mahalanobis foi determinada utilizando o programa GENES versão 2009.7.0 (CRUZ, 2006).

3.2.2. Método da variância mínima de Ward

O método de agrupamento escolhido para a análise foi o Método de Ward. Esse método, proposto por Ward (1963) baseia-se na mudança de variação dentro dos grupos em formação e entre eles a cada passo do processo de agrupamento. O procedimento também é conhecido como “método da variância mínima”, pois considera como função objetivo minimizar a soma de quadrados dentro dos grupos.

Para a formação inicial do grupo esse método considera o par de indivíduos que proporciona a menor soma de quadrados dos desvios. Admite-se que, em qualquer estágio, há perda de informações em razão do agrupamento realizado, o qual pode ser quantificado pela razão entre a soma de quadrados dos desvios dentro do grupo em formação pela soma de quadrado total dos desvios. A soma de quadrados dos desvios dentro é calculada considerando apenas os acessos dentro do grupo em formação e a soma de quadrados dos desvios total é calculada considerando todos os indivíduos disponíveis para análise de agrupamento.

O agrupamento é feito a partir das somas de quadrados dos desvios entre acessos ou, alternativamente, a partir do quadrado da distância Euclidiana, uma vez que se verifica a relação:

$$SQD_{ii'} = \frac{1}{2} d_{ii'}^2$$

em que:

$$SQD_{ii'} = \sum_{j=1}^V SQD_{j(ii')}$$

sendo $SQD_{j(ii')}$ a soma de quadrados dos desvios, para a j -ésima variável, considerando os acessos i e i' .

$$d_{ii'}^2 = \sum_{j=1}^V (Y_{ij} - Y_{i'j})^2$$

em que:

$d_{ii'}^2$: quadrado da distância euclidiana entre os acessos i e i' ;

V : número de características avaliadas;

Y_{ij} : valor da característica j para o acesso i .

A soma de quadrados dos desvios total é dada por:

$$SQDTotal = \frac{1}{g} \sum_{i < i'}^g \sum_{i''}^g d_{ii''}^2$$

sendo g o número de acessos.

Identifica-se então na matriz D (cujos elementos são os quadrados das distâncias Euclidianas - $d_{ii''}^2$) ou na matriz S (cujos elementos são as somas dos quadrados dos desvios - $SQD_{ii''}$) o par de acessos que proporciona menor soma de quadrados dos desvios. Com estes acessos agrupados, uma nova matriz de dissimilaridade, de dimensão inferior, é recalculada, considerando que:

$$SQD_{(ijk)} = \frac{1}{k} d_{(ijk)}^2$$

(k é o número de acessos no grupo, que neste caso é igual a 3).

$$d_{(ijk)}^2 = d_{(ij)}^2 + d_{(ij)k}^2 = d_{ij}^2 + d_{ik}^2 + d_{jk}^2$$

e ainda que:

$$SQD_{(ijkm)} = \frac{1}{k} d_{(ijkm)}^2$$

(k é o número de acessos no grupo, que neste caso é igual a 4).

$$d_{(ijkm)}^2 = d_{ij}^2 + d_{ik}^2 + d_{jk}^2 + d_{im}^2 + d_{jm}^2 + d_{km}^2$$

e assim sucessivamente.

No procedimento realiza-se a análise de agrupamento fornecendo os $g-1$ (g é o número de acessos a serem agrupados) passos de agrupamento para que seja formado o dendrograma.

Como o Método de Ward (1963) originalmente tem por base a distância euclidiana e nesse trabalho utilizou como medida de dissimilaridade a Distância generalizada de Mahalanobis, foi considerado apenas o algoritmo do método.

O Método de Ward pode ser definido e implementado pela família de algoritmos Lance e Williams (1967), que propuseram um método bastante geral para atualização da matriz de distâncias. Ferreira (2008) salienta que o método engloba todos os outros métodos hierárquicos aglomerativos, bastando para isso determinar adequadamente os parâmetros $(\alpha, \beta, \Upsilon)$ da fórmula geral proposta para atualizar a distância entre as entidades i e j , recém agrupadas, e uma entidade k . A fórmula de atualização é dada por:

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

em que, para o algoritmo de Ward:

$$\alpha_i = \frac{n_i + n_k}{n_i + n_j + n_k}, \alpha_j = \frac{n_j + n_k}{n_i + n_j + n_k}, \alpha_k = \frac{-n_k}{n_i + n_j + n_k}, \Upsilon = 0$$

Nesse caso, $d_{(ij)k}$ representa a distância entre as entidades i e j , recentemente agrupadas, e a entidade k ; n_i , n_j e n_k representam o número de indivíduos em cada entidade.

3.2.3. Determinação do número de grupos

Para determinar o número de grupos na análise de agrupamento, utilizou-se nesse trabalho o procedimento sugerido por Mojena (1977). Este se baseia no tamanho relativo dos níveis de fusões (distâncias) no dendrograma para determinar o ponto de corte dos dendrogramas gerados pelos métodos hierárquicos e, assim, definir o número de grupos. Propõe

selecionar o número de grupos no passo j que, primeiramente, satisfizer a seguinte inequação:

$$\alpha_{j+1} > \theta_k$$

em que α_{j+1} é o valor de distância do nível de fusão correspondente ao passo $j+1$ 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 a ser adotado, conforme sugerido por Milligan e Cooper (1985), é de $k=1,25$ como regra de parada na definição do número de grupos.

Assim, tem-se que:

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

A análise de agrupamento foi realizada utilizando a função *hclust*, pertencente ao pacote *stats* do software livre R versão 2.13.1 (R DEVELOPMENT CORE TEAM, 2011).

3.3. Modelos de regressão não linear

Assim que os grupos de acessos semelhantes foram definidos, ajustou-se sete modelos de regressão não linear apresentados na Tabela 2 às médias de matéria seca do bulbo (MSB), da raiz (MSR) e total da planta

(MSTP) de cada grupo, utilizando como variável preditora os dias após plantio.

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

Modelo	Modelo Estatístico	Referência
A	$y_i = \beta_1 \left(1 - e^{(\beta_3 \beta_2 - \beta_3 x_i)} \right) + \varepsilon_i$	Mitscherlich (1919)
B	$y_i = \beta_1 e^{\left(-\beta_2 e^{(-\beta_3 x_i)} \right)} + \varepsilon_i$	Gompertz (GOMPERTZ e PHILOS, 1825)
C	$y_i = \frac{\beta_1}{1 + \beta_2 e^{(-\beta_3 x_i)}} + \varepsilon_i$	Logístico (RATKOWSKI, 1983)
D	$y_i = \beta_1 - \beta_2 e^{(-\beta_3 x_i)} + \varepsilon_i$	Meloun I (MELOUN e MILITKY, 1996)
E	$y_i = \beta_1 - e^{(-\beta_2 - \beta_3 x_i)} + \varepsilon_i$	Meloun II (MELOUN e MILITKY, 1996)
F	$y_i = \beta_1 \left(1 - \beta_2 e^{(-\beta_3 x_i)} \right) + \varepsilon_i$	Brody (1945)
G	$y_i = \beta_1 \left(1 - \beta_2 e^{(-\beta_3 x_i)} \right)^3 + \varepsilon_i$	von Bertalanffy (1957)

em que:

β_1 é o parâmetro que representa o peso assintótico;

β_2 é um parâmetro de locação, sem interpretação biológica;

β_3 é o parâmetro que representa a taxa de maturidade, ou velocidade do crescimento;

y_i representa a observação na variável dependente (ou variável resposta);

x_i representa a variável independente (ou variável preditora);

ε_i representa o erro aleatório, $\varepsilon_i \sim N(0, \sigma^2)$.

Os modelos de Mistcherlich (A), Gompertz (B), Logístico (C), Brody (F) e von Bertalanffy (G) são bastante utilizados na literatura para descrever curvas de crescimento de animais e plantas e também acúmulo de matéria seca, em que Santoro et al. (2005), Sarmiento et al. (2006) e Maia et al. (2009) utilizam os modelos B, C, F e G; Pôrto et al. (2007) utilizam o modelo C; Brito et al. (2007) utilizam os modelos A e G; Silveira et al. (2011) utilizam os modelos A, B, C, F e G. Já os modelos Meloun I e Meloun II, propostos por Meloun e Militky (1996) aparecem em poucos trabalhos, pode-se citar Silveira et al. (2011) e Silva et al. (2011).

Para o ajuste dos modelos utilizou-se a função *nls*, pertencente ao pacote *stats* do software livre R versão 2.13.1 (R DEVELOPMENT CORE TEAM, 2011), que tem por base o método dos mínimos quadrados ordinários. O método iterativo utilizado foi o de Gauss-Newton.

3.3.1. Método dos mínimos quadrados ordinários

Seja o modelo (1) do item 2.3 escrito da seguinte forma matricial:

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$$

em que:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{f}(\boldsymbol{\theta}) = \begin{bmatrix} f(x_1, \boldsymbol{\theta}) \\ f(x_2, \boldsymbol{\theta}) \\ \vdots \\ f(x_n, \boldsymbol{\theta}) \end{bmatrix} \quad \text{e} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

A soma dos quadrados dos erros aleatórios (SSE) deverá ser minimizada por $\boldsymbol{\theta}$. Assim, função de quadrados mínimos para um modelo não linear é dada por:

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

ou, então em notação vetorial:

$$\begin{aligned} SSE(\boldsymbol{\theta}) &= \|\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})\|^2 \\ SSE(\boldsymbol{\theta}) &= [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})]'[\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})] \end{aligned} \quad (2)$$

em que $\|\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})\|$ é a norma euclidiana de $\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})$.

Visando um melhor entendimento desse processo de estimação, faz-se necessário definir as seguintes notações matriciais:

$$\mathbf{f}(\boldsymbol{\theta}) = \begin{bmatrix} f_1(\boldsymbol{\theta}) \\ f_2(\boldsymbol{\theta}) \\ \vdots \\ f_n(\boldsymbol{\theta}) \end{bmatrix}$$

e

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

em que $\mathbf{f}(\boldsymbol{\theta})$ é uma função vetor de dimensão $n \times 1$ do vetor de parâmetros $\boldsymbol{\theta}$, de dimensão $p \times 1$ e $\mathbf{F}(\boldsymbol{\theta})$ é a matriz Jacobiana de $\mathbf{f}(\boldsymbol{\theta})$.

O estimador de mínimos quadrados $\hat{\boldsymbol{\theta}}$ então deve satisfazer a equação:

$$\left. \frac{\partial SSE(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}' = \hat{\boldsymbol{\theta}}'} = \mathbf{0}$$

sendo

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

Assim,

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

Portanto, o sistema equações normais (SEN) para regressão não linear é dado por:

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

Estas equações são não lineares em relação aos estimadores dos parâmetros e, em geral, não apresentam fórmulas explícitas, fazendo necessários processos iterativos para obter as estimativas dos parâmetros.

3.3.2. Processo iterativo de Gauss-Newton

Um método amplamente usado em algoritmos computacionais para regressão não linear é a linearização da função não linear conhecido por método iterativo de Gauss-Newton (REGAZZI, 2003). O argumento básico desse método é como segue.

Suponha que se conheça uma aproximação inicial $\hat{\boldsymbol{\theta}}_0$ para $\hat{\boldsymbol{\theta}}$. As expansões de Taylor:

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

das componentes de $\mathbf{f}(\boldsymbol{\theta})$ na vizinhança de $\hat{\boldsymbol{\theta}}_0$ produzem a aproximação linear:

$$\mathbf{f}(\boldsymbol{\theta}) \cong \mathbf{f}(\hat{\boldsymbol{\theta}}_0) + \mathbf{F}(\hat{\boldsymbol{\theta}}_0)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0) \quad (3)$$

em que:

$$\mathbf{F}(\hat{\boldsymbol{\theta}}_0) = \left. \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}$$

Assim, o problema de minimizar a função (2) no item 3.3.1, passa a ser o de minimizar a função $SSE(\boldsymbol{\theta})$ associada à aproximação (3), dada por:

$$SSE(\boldsymbol{\theta}) = \|\mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\theta}}_0) - \mathbf{F}(\hat{\boldsymbol{\theta}}_0)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0)\|^2 \quad (4)$$

Fazendo $\mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\theta}}_0) = \mathbf{E}_0$ e $\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0 = \Delta\hat{\boldsymbol{\theta}}_0$ em (4), temos que:

$$\begin{aligned} SSE(\boldsymbol{\theta}) &= \|\mathbf{E}_0 - \mathbf{F}(\hat{\boldsymbol{\theta}}_0)\Delta\hat{\boldsymbol{\theta}}_0\|^2 \\ SSE(\boldsymbol{\theta}) &= [\mathbf{E}_0 - \mathbf{F}(\hat{\boldsymbol{\theta}}_0)\Delta\hat{\boldsymbol{\theta}}_0]' [\mathbf{E}_0 - \mathbf{F}(\hat{\boldsymbol{\theta}}_0)\Delta\hat{\boldsymbol{\theta}}_0] \\ SSE(\boldsymbol{\theta}) &= \mathbf{E}_0' \mathbf{E}_0 - \mathbf{E}_0' \mathbf{F}(\hat{\boldsymbol{\theta}}_0) \Delta\hat{\boldsymbol{\theta}}_0 - \Delta\hat{\boldsymbol{\theta}}_0' \mathbf{F}'(\hat{\boldsymbol{\theta}}_0) \mathbf{E}_0 + \Delta\hat{\boldsymbol{\theta}}_0' \mathbf{F}'(\hat{\boldsymbol{\theta}}_0) \mathbf{F}(\hat{\boldsymbol{\theta}}_0) \Delta\hat{\boldsymbol{\theta}}_0 \\ SSE(\boldsymbol{\theta}) &= \mathbf{E}_0' \mathbf{E}_0 - 2\Delta\hat{\boldsymbol{\theta}}_0' \mathbf{F}'(\hat{\boldsymbol{\theta}}_0) \mathbf{E}_0 + \Delta\hat{\boldsymbol{\theta}}_0' \mathbf{F}'(\hat{\boldsymbol{\theta}}_0) \mathbf{F}(\hat{\boldsymbol{\theta}}_0) \Delta\hat{\boldsymbol{\theta}}_0 \end{aligned}$$

Então a derivada de $SSE(\boldsymbol{\theta})$ em relação a $\Delta\hat{\boldsymbol{\theta}}_0$ é dada por:

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

Fazendo então:

$$\frac{\partial \text{SSE}(\boldsymbol{\theta})}{\partial \Delta \hat{\boldsymbol{\theta}}_0} = \mathbf{0}$$

obtemos o Sistema de equações normais (SEN):

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

em que $\mathbf{E}_0 = \mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\theta}}_0)$ desempenha o papel de variável dependente $\mathbf{y}$ e $\mathbf{F}'(\hat{\boldsymbol{\theta}}_0)$ o da matriz X dos modelos lineares.

Se $\mathbf{F}(\hat{\boldsymbol{\theta}}_0)$ apresentar posto coluna completo, o valor de $\Delta \hat{\boldsymbol{\theta}}_0$ que minimiza $\text{SSE}(\boldsymbol{\theta})$ é:

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

Uma vez que $\Delta \hat{\boldsymbol{\theta}}_0 = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0$, pode-se definir um vetor:

$$\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}}_0 + \Delta \hat{\boldsymbol{\theta}}_0$$

como estimativas atualizadas de $\boldsymbol{\theta}$.

Utilizando o mesmo procedimento, desta vez com as estimativas atualizadas $\hat{\boldsymbol{\theta}}_1$ no lugar de $\hat{\boldsymbol{\theta}}_0$, será produzido outro conjunto de estimativas atualizadas $\hat{\boldsymbol{\theta}}_2$, e assim por diante.

Em geral, para a k-ésima iteração tem-se:

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

em que

$$\mathbf{F}(\hat{\boldsymbol{\theta}}_k) = \left. \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_k}$$

$$\mathbf{E}_k = \mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\theta}}_k)$$

$$\hat{\boldsymbol{\theta}}_k = [\hat{\theta}_{1k} \quad \hat{\theta}_{2k} \quad \dots \quad \hat{\theta}_{pk}]'$$

Esse processo iterativo continua até a convergência, ou seja, até:

$$\left| \frac{\hat{\theta}_{j(k+1)} - \hat{\theta}_{jk}}{\hat{\theta}_{jk}} \right| < \delta \quad j = 1, 2, \dots, p$$

em que δ é algum número pequeno, no *default* da função utilizada no software R, *nls*, esse valor é $\delta = 1 \times 10^{-5}$. Pode-se avaliar $SSE(\hat{\boldsymbol{\theta}}_k)$ a cada iteração para garantir que foi obtida uma redução neste valor.

Os valores iniciais foram definidos a partir de uma transformação que "lineariza" a função, onde os dados transformados foram submetidos à regressão linear simples e as estimativas dos parâmetros desse ajuste foram utilizadas, de acordo com a transformação de cada modelo, no cálculo dos valores iniciais.

E, segundo Souza (1998), em modelos de regressão não linear não se pode fazer afirmações gerais sobre as propriedades dos estimadores de quadrados mínimos, tais como não tendenciosidade e variância mínima, exceto para grandes amostras.

3.3.3. Avaliadores da qualidade do ajuste

Depois de ajustados os modelos, calculou-se três avaliadores de qualidade de ajuste para cada modelo. Os avaliadores calculados foram o coeficiente de determinação (R^2), o quadrado médio do resíduo (QMR) e o desvio médio absoluto dos resíduos (DMA), proposto por Sarmento et al. (2006).

3.3.3.1. Coeficiente de determinação (R^2)

Há algumas discussões sobre o R^2 para regressão não linear, por exemplo, Ratkowsky (1990) comenta que o R^2 não tem nenhum significado óbvio no caso de modelos de regressão não linear e não precisa ser calculado.

Já Souza (1998) observa que a qualidade do ajustamento pode ser medida pelo quadrado do coeficiente de correlação entre os valores observados e preditos, mesmo no caso de modelo de regressão não linear, que pode ser calculada também com a utilização da fórmula dada a seguir:

$$R^2 = 1 - \frac{SQR}{SQT} \quad (5)$$

em que SQR é a soma dos quadrados residuais e SQT é a soma de quadrados total corrigida pela média.

Quanto maior é o coeficiente de determinação, melhor é o ajuste dos dados ao modelo.

Nesse trabalho, assim como em outros encontrados na literatura (AGOSTINETTO et al., 2004; SARMENTO et al., 2006; MAIA et al., 2009), foi utilizada a equação (5) para o cálculo do coeficiente de determinação.

3.3.3.2. Quadrado médio do resíduo (QMR)

O Quadrado médio do resíduo 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 da característica em questão;

$\hat{y}_i$ é o valor estimado matéria seca da característica em questão pelo modelo analisado.

Quanto menor é o valor de QMR , melhor é o ajuste dos dados ao modelo.

3.3.3.3. Desvio médio absoluto dos resíduos (DMA)

Esse avaliador da qualidade do ajuste foi proposto por Sarmento et al. (2006), onde o desvio médio absoluto dos resíduos é dado por:

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

em que n é o número de observações, y_i é o valor observado da matéria seca da característica em questão e $\hat{y}_i$ é o valor estimado matéria seca da característica em questão pelo modelo analisado.

Quanto menor o desvio médio absoluto dos resíduos, melhor é o ajuste dos dados ao modelo.

Após calculados os avaliadores da qualidade do ajuste analisou-se quais dos modelos referentes ao acúmulo de cada matéria seca de cada grupo melhor se ajustaram.

4. RESULTADOS E DISCUSSÃO

4.1. Agrupamento dos acessos

Utilizando o método de Mojena (1977) e $k = 1,25$ determinou-se um ponto de corte no dendrograma obtido com o algoritmo de Ward na distância 16,63 (valor do critério θ_k do método de Mojena), que corresponde a 40,06% da distância máxima observada nos níveis de fusão (Figura 1).

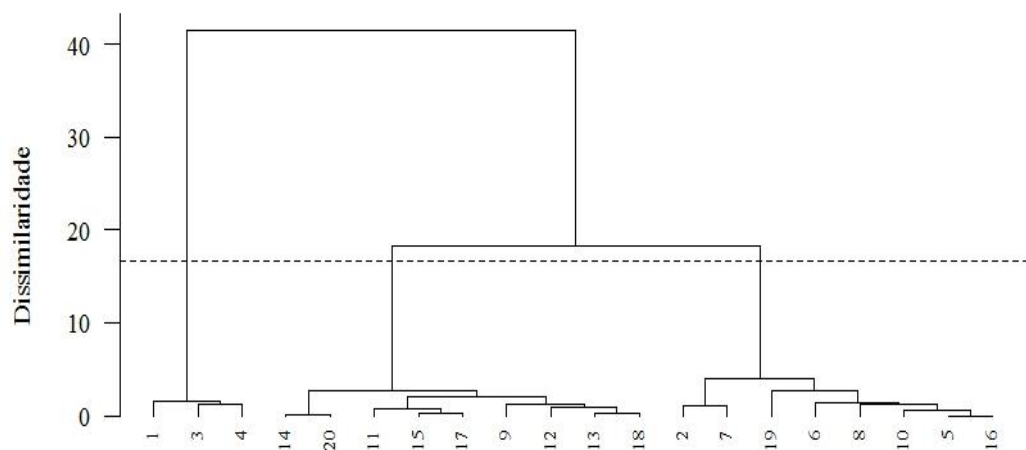


Figura 1. Dendrograma obtido com o algoritmo de Ward com base na distância generalizada de Mahalanobis (D^2) para agrupamento de 20 acessos de alho pertencentes ao BGH/UFV. UFV, Viçosa (MG), 2012

Assim, foram definidos três grupos de acessos, como se observa na Tabela 3.

Tabela 3. Formação dos grupos de 20 acessos de alho do BGH/UFV obtidos com o algoritmo de Ward com base na distância generalizada de Mahalanobis (D^2). UFV, Viçosa (MG), 2012.

Grupo	Nº de acessos		Acessos							
I	3				1	3	4			
II	9	9	11	12	13	14	15	17	18	20
III	8		2	5	6	7	8	10	16	19

As médias de cada grupo para a matéria seca do bulbo (MSB), da raiz (MSR) e total da planta (MSTP) em cada período avaliado estão apresentadas na Tabela 4.

Tabela 4. Valores médios das características de matéria seca do bulbo (MSB), da raiz (MSR) e total da planta (MSTP), em gramas por planta, dos respectivos grupos e dias após plantio. UFV, Viçosa (MG), 2012.

Características	Grupos	Dias após plantio			
		60	90	120	150
MSB	I	0,089	0,495	4,671	18,550
	II	0,098	0,688	6,783	14,117
	III	0,091	0,739	8,666	14,797
MSR	I	0,130	0,495	1,488	2,233
	II	0,093	0,274	0,713	1,106
	III	0,055	0,276	1,027	1,192
MSTP	I	0,996	4,558	19,083	28,258
	II	0,868	4,081	17,119	20,367
	III	0,683	3,902	17,126	20,003

Verifica-se que, para a matéria seca do bulbo, apesar do grupo I ter apresentado as menores médias nos três primeiros períodos avaliados, aos 150 dias após plantio a média dessa característica apresentou-se bem maior que as médias dos demais grupos. Já para as outras características, as matérias secas do grupo I se apresentaram mais altas em todos os períodos avaliados durante o crescimento da planta, indicando que os acessos do grupo I, no geral, possuem maiores matérias secas das três características (bulbo, raiz e planta toda).

4.2. Estimativas dos parâmetros dos modelos de regressão não linear e avaliadores da qualidade do ajuste

Os dados contidos na Tabela 4 do item 4.1 então foram submetidos ao ajuste dos sete modelos apresentados na Tabela 2 do item 3.3. Para todos os grupos houve convergência de pelo menos dois modelos. Nas Tabelas 5, 6 e 7, são apresentadas as estimativas dos parâmetros e os avaliadores da qualidade do ajuste, de acordo com os modelos ajustados, para as médias de matéria seca do bulbo (MSB), da raiz (MSR) e total da planta (MSTP) de cada grupo de acesso, respectivamente.

Observa-se na Tabela 5 que o coeficiente de determinação R^2 foi de 0,999 para todos os modelos ajustados nos dois primeiros grupos e no grupo III ele só foi diferente em um modelo, o de von Bertalanffy, ainda assim com valores bem próximos de 1, indicando que, de modo geral, os dados dos três grupos se ajustaram bem aos modelos que convergiram. Comparando os valores de quadrado médio do resíduo (QMR) e de desvio médio absoluto dos resíduos (DMA) dos modelos ajustados dentro de cada grupo de acessos, pode-se perceber que, apesar da pequena diferença entre eles, o modelo Logístico foi o melhor ajustado, mostrando os menores valores para esses dois últimos avaliadores. Verifica-se ainda que essas diferenças entre o modelo Logístico e de Gompertz foram menores que as diferenças entre o modelo Logístico e de von Bertalanffy.

Pode-se verificar também que, no grupo I, o modelo Logístico indicou um peso assintótico de 26,323 g, já o modelo de Gompertz apresentou

Tabela 5. Estimativas dos parâmetros β_1 , β_2 e β_3 , coeficiente de determinação (R^2), quadrado médio do resíduo (QMR) e desvio médio absoluto dos resíduos (DMA), de acordo com os modelos estudados para as médias de matéria seca do bulbo (MSB) de cada grupo de acesso. UFV, Viçosa (MG), 2012.

Grupos	Modelos	Estimativas			Avaliadores		
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	R^2	QMR	DMA
I	B – Gompertz	175,167	24,618	0,016	0,999	0,006	0,021
	C - Logístico	26,323	69475,178	0,080	0,999	0,002	0,013
II	B – Gompertz	19,945	102,321	0,038	0,999	0,009	0,025
	C - Logístico	15,151	99438,967	0,094	0,999	0,003	0,016
	G – von Bertalanffy	33,187	3,248	0,017	0,999	0,141	0,160
III	B – Gompertz	17,167	306,336	0,051	0,999	0,008	0,023
	C - Logístico	15,234	329609,223	0,108	0,999	0,004	0,017
	G – von Bertalanffy	24,000	5,112	0,024	0,994	0,831	0,404

$\hat{\beta}_1 = 175,167$, valor esse alto demais para representar peso assintótico da matéria seca do bulbo, não se mostrando um bom representante para essa característica. Já nos dois outros grupos, as estimativas para o parâmetro $\hat{\beta}_1$, que representa o peso assintótico da matéria seca do bulbo, foram mais aproximadas entre si, sem apresentar uma dispersão maior como no grupo I.

As estimativas $\hat{\beta}_2$ dos modelos Logísticos ajustados nos três grupos de acessos foram relativamente altas, porém, como este não tem interpretação biológica, sendo apenas um parâmetro de locação, essa ocorrência não indicou nenhum problema de ajuste.

A Figura 2 apresenta as curvas dos modelos que tiveram os melhores ajustes para cada grupo de acessos em relação à matéria seca do bulbo que, como comentado acima, foi o modelo Logístico.

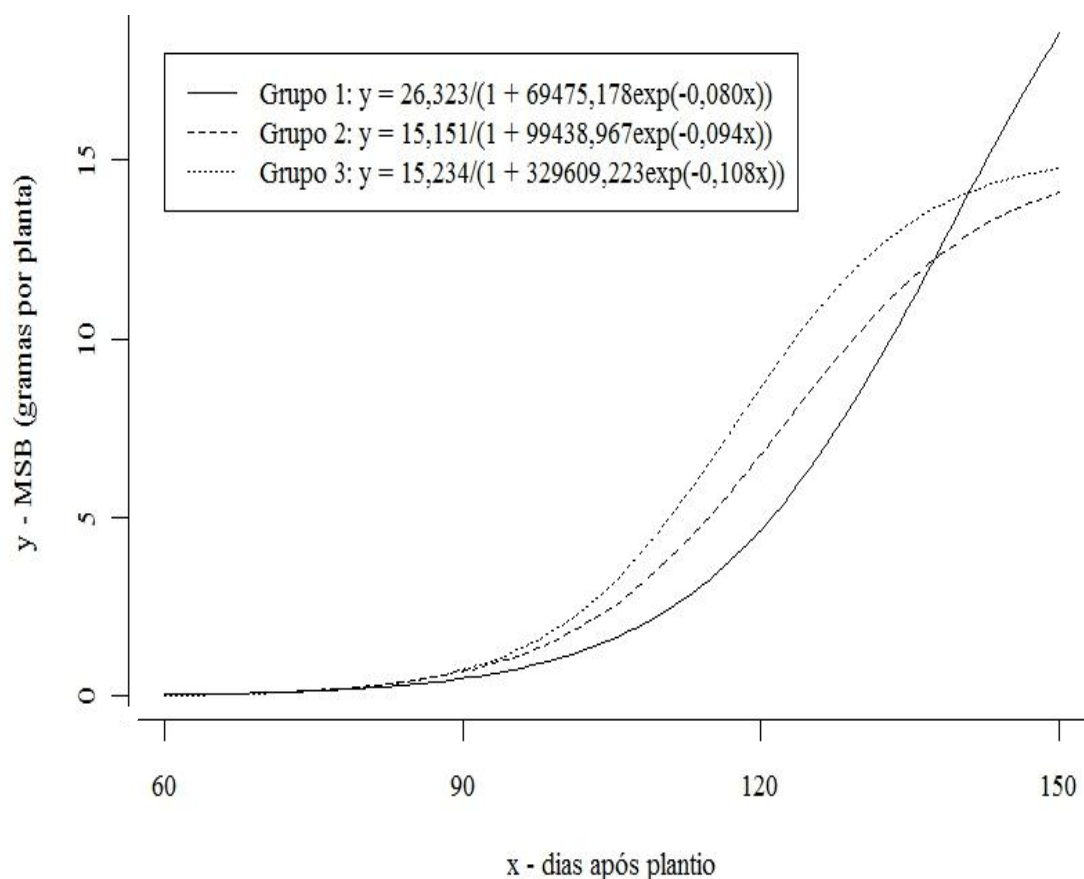


Figura 2. Curvas de acúmulo de matéria seca do bulbo (MSB) ajustadas pelo modelo Logístico dos três grupos de acessos de alho. UFV, Viçosa (MG), 2012.

Pode-se verificar nesse gráfico que a curva que representa o melhor modelo ajustado para a matéria seca do bulbo (MSB) no grupo I, o Logístico, estava abaixo das curvas do modelo Logístico dos outros dois grupos, ultrapassando-as por volta dos 140 dias após plantio. Essa ocorrência pode indicar ao pesquisador, se houver pretensão de colheita antes de 140 dias após o plantio, uma recomendação de que se utilize acessos semelhantes aos acessos que formam o grupo 3, que se mostrou superior até esse período. A partir daí, a sugestão é para que se utilize acessos semelhantes aos pertencentes ao grupo 1.

Analisa-se também que os modelos Logísticos representantes dos grupos II e III apresentaram um formato sigmoidal bem mais definido do que o grupo I. Formato esse que se pode observar bem definidamente em trabalhos envolvendo curvas de crescimento de bananeiras e acúmulo de matéria seca do bulbo de cebolas (MAIA et al., 2009; PÔRTO et al., 2007).

A Tabela 6 apresenta que o coeficiente de determinação (R^2) variou de 0,995 a 0,999 no grupo I, de 0,996 a 0,999 no grupo II e de 0,933 a 0,999 no grupo III para o ajuste dos modelos referentes às médias de matéria seca da raiz (MSR), sendo que o maior valor (de 0,999) foi calculado para o modelo Logístico nos três grupos. Observa-se também que no grupo II, o QMR e DMA tiveram uma menor variação, com $0,001 \leq QMR \leq 0,003$ e $0,007 \leq DMA \leq 0,024$, e no grupo III, maior variação, com $0,001 \leq QMR \leq 0,062$ e $0,013 \leq DMA \leq 0,112$.

Todos os valores mínimos calculados para esses dois avaliadores foram provenientes do ajuste para o modelo Logístico, concordando com a análise do R^2 de que este modelo foi o melhor ajustado para matéria seca da raiz nos três grupos, já que esses dois últimos avaliadores são julgados como melhores quando seus valores são menores.

As curvas do modelo Logístico, que foi o modelo melhor para os três grupos de acessos em relação à matéria seca da raiz, seguem apresentadas na Figura 3.

As informações obtidas na Figura 3 condizem com os valores observados na Tabela 4 do item 4.1. Essa tabela aponta que as maiores médias de matéria seca da raiz foram apresentadas pelo grupo I e isto é verificado na Figura 3 quando se observa que a curva contínua, referente ao

Tabela 6. Estimativas dos parâmetros β_1 , β_2 e β_3 , coeficiente de determinação (R^2), quadrado médio do resíduo (QMR) e desvio médio absoluto dos resíduos (DMA), de acordo com os modelos estudados para as médias de matéria seca da raiz (MSR) de cada grupo de acesso. UFV, Viçosa (MG), 2012.

Grupos	Modelos	Estimativas			Avaliadores		
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	R^2	QMR	DMA
I	B – Gompertz	3,100	22,042	0,028	0,997	0,009	0,040
	C - Logístico	2,505	730,822	0,058	0,999	0,001	0,011
	G – von Bertalanffy	3,795	2,411	0,018	0,995	0,017	0,056
II	B – Gompertz	2,004	10,938	0,019	0,997	0,002	0,019
	C - Logístico	1,345	273,042	0,048	0,999	0,001	0,007
	G – von Bertalanffy	3,350	1,303	0,010	0,996	0,003	0,024
III	B – Gompertz	1,215	1004,136	0,072	0,997	0,003	0,014
	C - Logístico	1,208	19044,764	0,096	0,999	0,001	0,013
	F - Brody	8,649	1,109	0,002	0,933	0,062	0,112
	G – von Bertalanffy	1,367	13,106	0,040	0,985	0,016	0,061

grupo I, está sempre acima das demais curvas, representantes dos outros dois grupos de acessos, lembrando que o modelo Logístico foi o melhor ajustado também para essa característica.

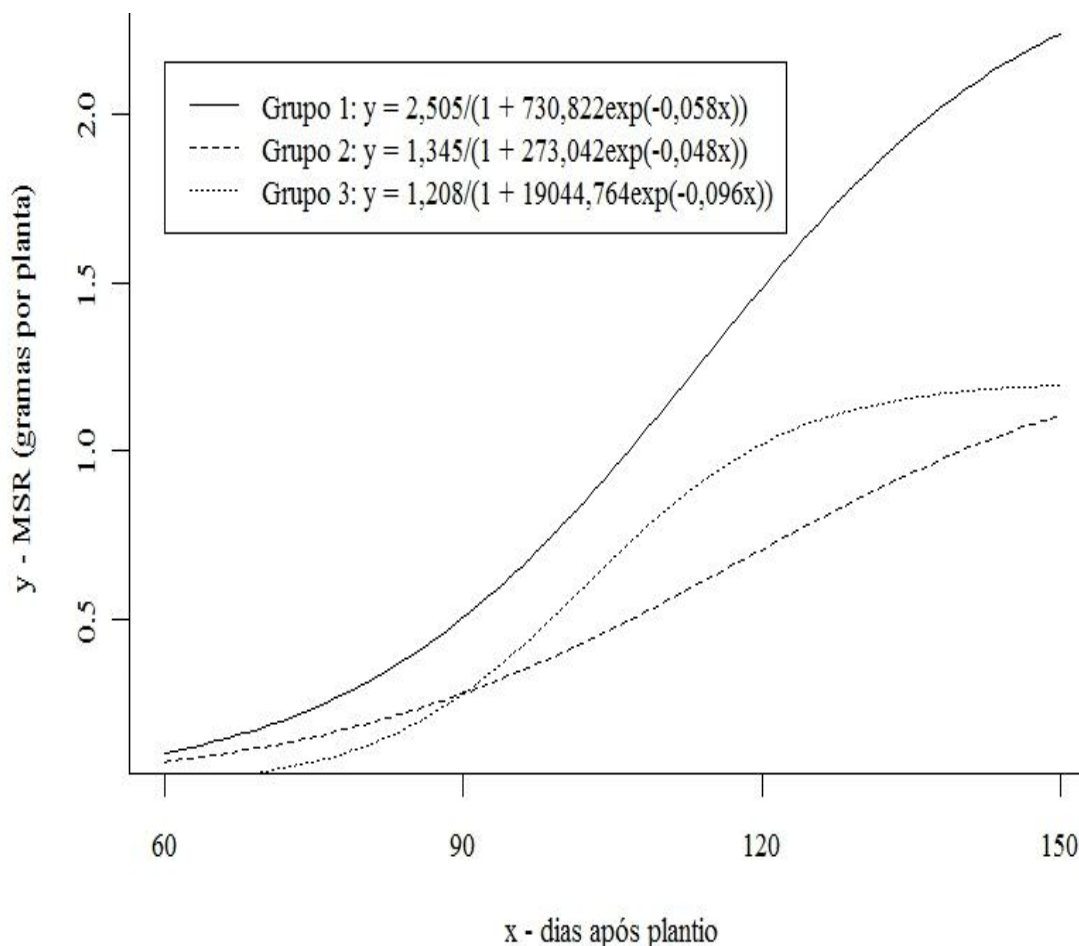


Figura 3. Curvas de acúmulo de matéria seca da raiz (MSR) ajustadas pelo modelo Logístico dos três grupos de acessos de alho. UFV, Viçosa (MG), 2012.

Verifica-se que o grupo II já atingiu seu peso assintótico por volta dos 130 dias após plantio, enquanto grupos I e III demoraram mais para atingir o peso assintótico, por volta dos 150 dias após plantio. (Figura 3).

É observado na Tabela 7 que todos os sete modelos propostos tiveram convergência para as médias de matéria seca total da planta (MSTP) referentes ao grupo III. Os avaliadores da qualidade do ajuste dessa característica nesse grupo tiveram variações de $0,924 \leq R^2 \leq 0,999$, $0,224 \leq QMR \leq 20,721$ e $0,150 \leq DMA \leq 2,036$, em que o maior valor de R^2

Tabela 7. Estimativas dos parâmetros β_1 , β_2 e β_3 , coeficiente de determinação (R^2), quadrado médio do resíduo (QMR) e desvio médio absoluto dos resíduos (DMA), de acordo com os modelos estudados para as médias de matéria seca total da planta (MSTP) de cada grupo de acesso. UFV, Viçosa (MG), 2012.

Grupos	Modelos	Estimativas			Avaliadores		
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	R^2	QMR	DMA
I	B – Gompertz	32,904	91,689	0,043	0,998	0,941	0,264
	C - Logístico	30,008	4472,638	0,075	0,999	0,191	0,150
	G – von Bertalanffy	37,414	6,360	0,029	0,997	1,673	0,542
II	B – Gompertz	20,861	917,014	0,070	0,998	0,753	0,217
	C - Logístico	20,633	26094,219	0,098	0,999	0,366	0,198
	G – von Bertalanffy	23,793	11,933	0,038	0,984	4,793	1,066
III	A - Mitscherlich	391,863	61,295	0,001	0,924	20,721	2,036
	B – Gompertz	20,371	1428,841	0,075	0,998	0,466	0,171
	C - Logístico	20,193	46565,013	0,104	0,999	0,224	0,150
	D – Meloun I	391,648	406,881	0,001	0,924	20,721	2,036
	E – Meloun II	391,346	-6,008	0,001	0,924	20,721	2,036
	F – Brody	391,857	1,039	0,001	0,924	20,721	2,036
	G – von Bertalanffy	23,340	12,655	0,039	0,982	5,352	1,125

e os menores valores de QMR e DMA foram observados quando se ajustou o modelo Logístico, indicando que o modelo Logístico também foi o melhor ajustado para a matéria seca total da planta no grupo III. Ainda em relação aos ajustes de modelos desse grupo, pode-se observar que as estimativas dos parâmetros β_1 e β_3 para os modelos de Mitscherlich, Meloun I, Meloun II e Brody são aproximadamente iguais, isso pode ser que ocorra devido ao fato de que esses quatro modelos são apenas reparametrizações diferentes de um mesmo modelo, em que as mudanças nos valores das estimativas só acontece para o parâmetro β_2 .

Os grupos I e II também apresentaram melhores avaliadores da qualidade do ajuste para o modelo Logístico, podendo dizer então que o modelo Logístico obteve melhor ajuste para os três grupos em relação à matéria seca total da planta. Assim, as curvas desse modelo para os três grupos de acessos em relação à matéria seca total da planta, seguem apresentadas na Figura 4.

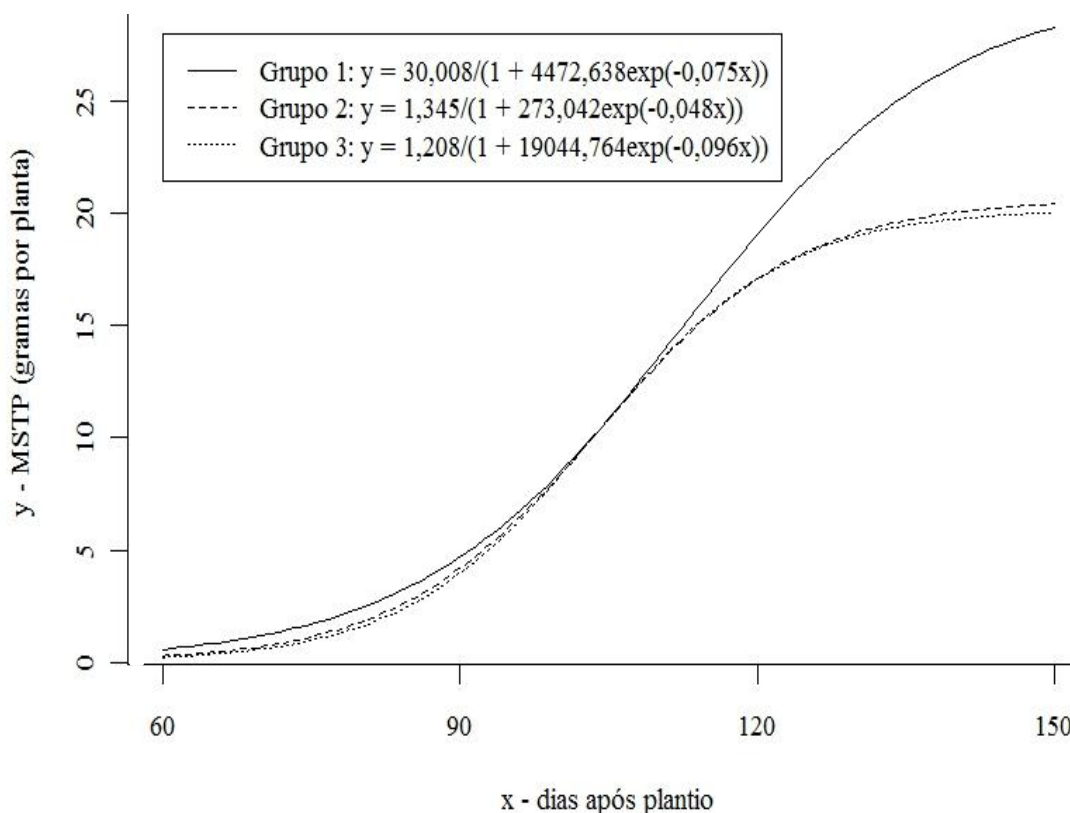


Figura 4. Curvas de acúmulo de matéria seca total da planta (MSTP) ajustadas pelo modelo Logístico dos três grupos de acessos de alho. UFV, Viçosa (MG), 2012.

Observa-se na Figura 4 que as curvas de modelo Logístico representantes dos grupos II e III para a matéria seca total da planta foram bem semelhantes, já o grupo I, que teve médias maiores em relação aos outros dois grupos, como foi verificado na Tabela 4 do item 4.1, apresentou também a curva do modelo Logístico acima das demais curvas. Além disso, verifica-se que as curvas dos grupos II e III atingiram seu peso assintótico por volta dos 130 dias após plantio, enquanto a curva o grupo I começou a se estabilizar já perto dos 150 dias.

No trabalho de Maia et al. (2009) verificou-se também que o modelo Logístico foi o melhor ajustado para descrever curvas de crescimento em bananeiras, além disso Pôrto et al. (2007), utiliza o modelo Logístico para descrever acúmulos de matéria seca da parte aérea, do bulbo e da planta inteira da cultura de cebola, tendo como resultados ótimos ajustes.

5. CONCLUSÕES

A análise de agrupamento realizada forneceu três grupos de acessos mais semelhantes, com três, oito e nove acessos cada um.

O modelo Logístico apresentou os melhores ajustes para descrever o acúmulo de matéria seca do bulbo, da raiz e total da planta de três grupos de acessos de alho pertencentes ao Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa, formados por meio de análise de agrupamento.

6. REFERÊNCIAS

AGOSTINETTO, D.; FLECK, N. G.; RIZZARDI, M. A.; BIANCHI, M. A.; SILVA, P. R. F. Comparação de modelos matemáticos na estimativa das perdas de produtividade de grãos em arroz irrigado. **Pesquisa Agropecuária Brasileira**, Brasília, v.39, n.9, p.847-856, 2004.

ALVAREZ, I.A.; CASTRO, P.R.C. Crescimento da parte aérea de cana crua e queimada. **Scientia Agricola**, v.56, p.1069-1079, 1999.

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

BRITO, C. C. R.; SILVA, J. A. A.; FERREIRA, R. L. C.; SANTOS, E. S.; FERRAZ, I. Modelos de crescimento resultantes da combinação e variações dos modelos de chapman-richards e silva-bailey aplicados em *Leucaena leucocephala* (Lam.) de Wit. **Ciência Florestal**, Santa Maria, v. 17, n. 2, p. 175-185, 2007.

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

CARGNELUTTI FILHO, A.; RIBEIRO, N. D.; BURIN, C. Consistência do padrão de agrupamento de cultivares de feijão conforme medidas de

dissimilaridade e métodos de agrupamento. **Pesquisa Agropecuária Brasileira**. Brasília, v.45, n.3, p.236-243, 2010.

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.

CECON, P. R.; SILVA, A. R.; NASCIMENTO, M.; FERREIRA, A. **Métodos Estatísticos**. Viçosa: UFV, 229p. 2012.

CRUZ, C.D. **Programa Genes: Biometria**. Editora UFV. Viçosa (MG). 382p. 2006.

CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. v. 2, 2 ed., Viçosa: UFV, 623p. 2003.

CRUZ, C. D.; FERREIRA, F. M.; PESSONI, L. A. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco: Suprema, 620p. 2011.

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

DINIZ FILHO, J. A. **Métodos filogenéticos comparativos**. Ribeirão Preto: Holos, 120p. 2000.

FARIA, P. N. **Avaliação de métodos para determinação do número ótimo de *clusters* em estudo de divergência genética entre acessos de pimenta**. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa. Orientador: Paulo Roberto Cecon, Viçosa, 67f. 2009.

FEITOSA, H. O.; JUNQUEIRA, R. M.; GUERRA, J. G. M.; TEIXEIRA, M. G.; RESENDE, F. V. Avaliação de cultivares de alho em três regiões do Estado do Rio de Janeiro cultivados sob sistema orgânico. **Revista Brasileira de Ciencias Agrárias**, v. 4, n. 4, p. 399-404, 2009.

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

FONSECA, A. F. A.; SEDIYAMA, T.; CRUZ, C. D.; SAKAIYAMA, N. S.; FERRÃO, M. A. G.; FERRÃO, R. G.; BRAGANÇA, S. M. Divergência genética em café conilon. **Pesquisa Agropecuária Brasileira**, v.41, n.4, p.599-605, 2006.

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

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Levantamento Sistemático da Produção Agrícola. Rio de Janeiro, v.25 n.05 p.1-88 maio.2012. Disponível em: http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/lspa/lspa_201205.pdf. Acesso em: 26, jun. 2012.

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

KHATTREE, R.; NAIK, D. N. **Multivariate data reduction and discrimination with SAS Software**. Cary, NC: SAS Institute Inc., 338p. 2000.

LANCE, G. N.; WILLIAMS, W. T. A General Theory of Classificatory Sorting Strategies. 1. Hierarchical Systems, **Computer Journal**, v.9, p.373-380, 1967.

MAHALANOBIS, P. C. On the generalized distance in statistics. **Proceedings of The National Institute of Sciences of India**, v.12, p.49-55, 1936.

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.

MANLY, B. F. J., **Multivariate Statistical Methods: A primer**. Second edition. London: Chapman & Hall, 215p.1994

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

MELOUN, M.; MILITKY, J. **Sbírka úloh, Statistické zpracování experimentálních dat**. Pardubice: Universita Pardubice. 1996.

MENEZES SOBRINHO, J.A. de; CHARCHAR, J.M.; ARAGÃO, F.A.S. Caracterização morfológica de germoplasma de alho por análises multivariada, componentes principais e variáveis canônicas. **Horticultura Brasileira**, Brasília, v. 17, n. 2, p. 96-101, 1999.

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.

OLIVEIRA, H.N.; LÔBO, R. B.; PEREIRA, C. S. Comparação de modelos não-lineares para descrever o crescimento de fêmeas da raça Guzerá. **Pesquisa Agropecuária Brasileira**, v.35, p.1843-1851, 2000.

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.

R DEVELOPMENT CORE TEAM. **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>. 2011

RATKOWSKY, D. A. **Handbook of nonlinear regressions models**. New York: Marcel Dekker Inc., 241 p. 1990.

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.

ROSEMBURG, H. C. **Cluster analysis for researchers**. California, EUA: Lifetime Learning, 334p. 1984.

SANTORO, K.R.; BARBOSA, B. P.; SANTOS, E. S.; BRASIL, L. H. A. Herdabilidade de parâmetros de curvas de crescimento não-lineares em

zebuínos, no estado de Pernambuco. **Revista Brasileira de Zootecnia**, v.34, p.2280-2289, 2005.

SARMENTO, J. L. R.; REGAZZI, A. J.; SOUZA, W. H.; TORRES, R. A.; BREDÁ, 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.

SAS Institute Inc. **Statistical analysis system user's guide**. Version 9, 1. ed. Cary: SAS Institute, USA, 2003.

SILVA, A. R., **Métodos de agrupamento: avaliação e aplicação ao Estudo de divergência genética em acessos de alho**. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa. Orientador: Paulo Roberto Cecon, Viçosa, 83f. 2012.

SILVA, F.F.; AQUINO, L. H.; OLIVEIRA, A.I.G. Influência de fatores genéticos e ambientais sobre as estimativas dos parâmetros das funções de crescimento em gado nelore. **Ciência e Agrotecnologia**, v. 25, n. 5, p.1195-1205, 2001.

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.

SILVEIRA, F. G. **Classificação multivariada de modelos de crescimento para grupos genéticos de ovinos de corte**. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa. Orientador: Fabyano Fonseca e Silva, Viçosa, 74f. 2010.

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

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

SOUZA, J.R.; SATURNINO, H.M.; MASCARENHAS, M.H.T.; LARA, J.F.R. Caracteres morfológicos de 18 cultivares de alho (*Allium sativum* L.). **Projeto Olericultura**; Relatório Anual 75/76, Belo Horizonte, p. 39- 43, 1977.

SOUZA, R. J.; MACÊDO, F. S.; YURI, J. E. Alho (*Allium sativum* L.). In: PAULA JÚNIOR, T.J.; VENZON, M. (Eds.). **101 Culturas**: Manual de tecnologias agrícolas. Belo Horizonte: EPAMIG, cap. 8, p.75-84. 2007.

TRANI, P.E. **Cultura do alho (*Allium sativum*): Diagnóstico e recomendações para seu cultivo no Estado de São Paulo**. 2009. Artigo em [Hypertexto](http://www.infobibos.com/Artigos/2009_2/alho/index.htm). Disponível em: http://www.infobibos.com/Artigos/2009_2/alho/index.htm. Acesso em: 29, jun. 2012.

UEDA, C. M.; YAMAMOTO, A. Y.; NUNES, W. M. C.; SCAPIM, C. A.; GUEDES, T. A. Nonlinear models for describing the Citrus Variegated Chlorosis in groves of two counties at Northwestern Paraná State, Brazil. **Acta Scientiarum. Agronomy**, v32, n. 4, p. 603-611, 2010.

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

WARD, J. Hierarchical grouping to optimize an objective function. **Journal of American Statistical Association**, v.58, p.236-244, 1963.

APÊNDICE

A – Script das análises de agrupamento no *Software R*

```
#####  
#                               Análise de Agrupamento  
#####  
  
#####  
# Distância generalizada de Mahalanobis  
#####  
  
mah=read.table("E:\\Dissertação\\Análises\\1. Agrupamento  
dos acessos\\mahalanobis",header=F)  
D2=as.matrix(mah)  
colnames(D2)=c(1:20)  
matrizD2=as.dist(m = D2, diag = FALSE, upper = FALSE)  
  
#####  
# Método de agrupamento de Ward  
#####  
  
k=1.25    # valor sugerido por Milligan e Cooper (1985)  
ward=hclust(matrizD2,method="ward")  
  
#####  
# Dendrograma e ponto de corte  
#####  
  
## Ponto de corte definido segundo Mojena (1977):  
  
PC=mean(ward$height)+k*sd(ward$height)
```

```
### Dendrograma:
```

```
par(family="serif",las=1,cex=1.2)
```

```
plot(ward, hang = -1,cex=0.7,main=" ",sub=" ", xlab=" ",ylab="Dissimilaridade")
```

```
abline(h=PC,v=NULL,col=1,lty=2)
```

B – Script dos ajustes dos modelos no *Software R*

```
# #####  
#           Ajustes dos modelos de regressão não linear  
# #####  
  
# #####  
# Definindo as variáveis  
# #####  
  
### Variável independente - Dia após plantio (DAP):  
  
x=c(60,90,120,150)  
  
### Variável dependente - médias da matéria seca total da  
### planta dos acessos pertencentes ao grupo 3, em gramas:  
  
y=c(0.683,3.902,17.126,20.003)  
  
# #####  
# Ajuste dos modelos  
# #####  
  
### Modelo A - Mitscherlich  
  
mA=nls(y~A*(1-exp(B*C-C*x)),start = list(A=10,B=100,C=0.1))  
  
summary(mA)  
  
### Modelo B - Gompertz  
  
mB=nls(y~A*exp(-B*exp(-C*x)),start = list(A=20,B=395,C=0.07))
```

```
summary(mB)
```

```
### Modelo C - Logístico
```

```
mC= nls(y ~ A/(1 + B*exp(-C*x)),  
start = list(A=20,B=395,C=0.07))
```

```
summary(mC)
```

```
### Modelo D - Meloun I
```

```
mD= nls(y~A-B*exp(-C*x), start = list(A=22,B=148,C=0.03),  
control=nls.control(maxiter=100))
```

```
summary(mD)
```

```
### Modelo E - Meloun II
```

```
mE= nls(y ~ A-exp(-B-C*x), start = list(A=22,B=-6,C=0.006),  
control=nls.control(maxiter=100,minFactor=1/2048))
```

```
summary(mE)
```

```
### Modelo F - Brody
```

```
mF= nls(y ~ A*(1-B*exp(-C*x)), start = list(A=20,B=51,C=0.05))
```

```
summary(mF)
```

```
### Modelo G - von Bertalanffy
```

```

mG= nls(y ~ A*(1-B*exp(-C*x))^3,
start = list(A=20,B=51,C=0.05))

summary(mG)

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

### Coeficiente de determinação (R2):

# soma de quadrados residual:
SQE_mC = summary(mC)$sigma^2*summary(mC)$df[2]

# soma de quadrado total corrigida:
SQT_mC = var(y)*(length(y)-1)

R2_mC = 1 - (SQE_mC)/(SQT_mC)

R2_mC

### Quadrado médio do resíduo (QMR):

# vetor dos valores estimados:

y_est=predict(mC,x)

QMR_mC=sum((y-y_est)^2)/summary(mC)$df[2])

QMR_mC

### Desvio médio absoluto do resíduo (DMA):

```

```
DMA_mC = sum(abs(y-y_est)/(length(y)))
```

```
DMA_mC
```

```
# #####  
# Plotar a curva  
# #####
```

```
par(family="serif",cex=0.9)
```

```
plot(x,y,ylab="MSTP (g)",xaxt="n",xlim=c(60,150))  
axis (1, xaxp = c(60,150,3))
```

```
curve(coef(mC)[1]/(1+coef(mC)[2]*exp(-  
coef(mC)[3]*x)),add=T,lty=1)
```

```
# #####  
# Observações  
# #####
```

```
### 1. Os valores iniciais foram definidos a partir de uma  
### transformação que "lineariza" a função, onde os dados  
### transformados foram submetidos a regressão linear simples  
### e as estimativas dos parâmetros desse ajuste foram  
### utilizadas, de alguma forma, no cálculo dos valores  
### iniciais.
```

```
### 2. O ajuste dos modelos para as outras características e  
### para os outros grupos, o ajuste é os cálculos dos  
### avaliadores para os outros modelos são realizados  
### utilizando os mesmos procedimentos acima.
```

```
### 3. A função nls ajusta modelos pelo método dos mínimos  
### quadrados ordinários e tem como default o método iterativo  
### de Gauss-Newton.
```