

ANÁLISE DE AGRUPAMENTO EM SELEÇÃO DE MODELOS DE REGRESSÃO NÃO LINEARES PARA DESCREVER O ACÚMULO DE MATÉRIA SECA EM PLANTAS DE ALHO

Guilherme Alves PUIATTI¹
Paulo Roberto CECON¹
Moysés NASCIMENTO¹
Mário PUIATTI²
Fernando Luiz FINGER²
Anderson Rodrigo da SILVA³
Ana Carolina Campana NASCIMENTO¹

- RESUMO: Este trabalho teve como objetivo 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. Vinte acessos de alho foram divididos em dois grupos, avaliados nas características: diâmetro, comprimento e peso médio do bulbo, número de bulbilhos e produtividade. Para os dois grupos foram ajustados modelos de regressão não lineares descrevendo o acúmulo de matéria seca da planta. Os modelos foram então agrupados segundo a qualidade de ajuste, medida pelo coeficiente de determinação, quadrado médio do resíduo, desvio médio absoluto dos resíduos, e os critérios de informação de Akaike e Bayesiano. A análise indicou o modelo logístico como o mais adequado para a descrição do acúmulo de matéria seca no alho, e um teste de identidade de modelos aplicado nos modelos dos dois grupos indicou que eles são estatisticamente iguais, e válidos para este estudo com alho.
- PALAVRAS-CHAVE: Classificação multivariada; identidade de modelos; *Allium sativum* L.

1 Introdução

O estudo da trajetória de crescimento das plantas é de grande importância para se fazer o manejo adequado destas. Esse conhecimento auxilia na preparação de técnicas de cultivo, colheita, conservação e também para a detecção de problemas no desenvolvimento de culturas. O interesse no estudo de crescimento de plantas aumentou nos últimos anos, conforme a evolução de técnicas computacionais mais ágeis e precisas para estas análises (Silva et al, 2001).

¹ Universidade Federal de Viçosa – UFV, Departamento de Estatística, CEP 36570-000, Viçosa, MG, Brasil. E-mail: guilherme@dpi.ufv.br / cecon@ufv.br / moysesnascim@ufv.br / ana.campana@ufv.br

² Universidade Federal de Viçosa – UFV, Departamento de Fitotecnia, CEP 36570-000, Viçosa, MG, Brasil. E-mail: mpuiatti@ufv.br / ffinger@ufv.br

³ Universidade de São Paulo – USP, Escola Superior de Agricultura Luiz de Queiroz – ESALQ, CEP 13418-900, Piracicaba, SP, Brasil. E-mail: anderson.agro@hotmail.com

O alho (*Allium sativum* L.) é uma planta condimentar 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. O bulbo é arredondado e pode conter de 5 a 56 bulbilhos. Estes têm a forma oval arqueada, sendo envolvidos por folhas protetoras (brácteas) de coloração branca ou arroxeadas. O alho possui características fortes de aroma e sabor que lhe atribuem propriedades condimentares que tornaram esta hortaliça destaque na culinária. Além disso, o alho possui grande valor nutricional, dado seu conteúdo em calorias, proteínas, carboidratos, fósforo, tiamina e vitamina B6 (Trani, 2009). Marchiori (2005) destaca que o alho apresenta diversas aplicações na medicina, principalmente por causa das suas características imuno estimulantes, antiarterioscleróticas, anticancerígenas e antimicrobianas.

De acordo com Honorato (2012), existe grande diversidade de cultivares de alho utilizados no Brasil; estas foram obtidas por meio de mutações somáticas e mediante seleção de características desejadas pelos próprios agricultores. Assim, visando auxiliar na preparação de técnicas de cultivo, colheita e de conservação, devido as diferenças existentes entre tais 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, e gera 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 exigindo diferentes quantidades de nutrientes.

Modelos de regressão não linear são os mais adequados para descrever estas curvas, uma vez que seus parâmetros possuem interpretação biológica, e processos como crescimento variam com o tempo (Mazzini et al, 2005). 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. 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, conforme as informações geradas pelos avaliadores de qualidade de ajuste, agrupar modelos mais próximos. 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 tais que exista grande homogeneidade dentro de cada grupo e heterogeneidade entre eles (Johnson; Wichern, 1992). Espera-se, assim, que modelos dentro de um mesmo grupo tenham uma interpretação e qualidade mais próximas dos dados em estudo.

No estudo de regressão de várias populações, com muita frequência, é de interesse saber quando um conjunto de equações ajustadas é idêntico, ou seja, se a relação de variáveis em estudo pode ser representada por uma única equação. Regazzi e Silva (2010) apresentam aproximações pelas estatísticas qui-quadrado e F para testar hipóteses de igualdade de subconjuntos de parâmetros e de identidade dos modelos para dados com repetições de experimento com delineamento em blocos completos casualizados. Um

mesmo modelo de regressão deve ser ajustado para todas as populações para que a técnica de identidade de modelos possa ser aplicada.

Diante do exposto, o objetivo deste trabalho foi utilizar análise de agrupamento para comparar e classificar modelos de regressão não lineares usados para descrever o acúmulo de matéria seca de plantas de alho ao longo do tempo, considerando os resultados dos avaliadores de qualidade de ajuste como variáveis. Dos modelos agrupados com melhores resultados, foi escolhido o modelo mais adequado a fim de se realizar um teste de identidade para comparar o ajuste do modelo nos diferentes grupos de acesso, e então verificar a possibilidade de uma equação comum para grupos de acessos com diferentes características.

2 Material e métodos

Os dados utilizados neste trabalho são provenientes de um experimento realizado no período de março a novembro de 2010, em área pertencente ao setor de olericultura do Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV), município de Viçosa, Zona da Mata de Minas Gerais. Foi utilizado o delineamento experimental de blocos completos casualizados com quatro repetições, cada bloco com duas plantas de cada acesso. Foram avaliados 81 acessos de alho registrados no Banco de Germoplasma de Hortaliças da UFV (BGH/UFV).

Para o estudo, foram escolhidos dez acessos de alho de dois grupos distintos, de forma que os acessos dentro de cada grupo apresentam características bem próximas. As características avaliadas foram: diâmetro do bulbo, comprimento do bulbo, peso médio de bulbo, número de bulbilhos por bulbo e produtividade. Os acessos selecionados são apresentados na Tabela 1.

Tabela 1 - Relação dos vinte acessos em estudo divididos em dois grupos, conforme seu registro no Banco de Germoplasmas de Hortaliças da Universidade Federal de Viçosa (BGH/UFV)

Grupo	Acessos				
1	0492	4483	4503	4825	4829
	4328	4491	4567	4826	4832
2	4484	4814	4816	4822	4838
	4493	4815	4820	4835	4844

2.1 Modelos de regressão não linear

Oito modelos de regressão não linear foram ajustados para cada grupo. Após a formação dos grupos, as médias da matéria seca total, medidas em gramas por planta, foram consideradas como variável resposta, enquanto a variável preditora é representada pelos dias após plantio (DAP), sendo avaliados os períodos de 60, 90, 120 e 150 dias após o plantio. A Tabela 2 apresenta os modelos ajustados.

Tabela 2 - Modelos de regressão não linear ajustados

Modelo	Função	Modelo	Função
A - Mitscherlich	$y_i = \beta_1(1 - e^{(\beta_2\beta_3 - \beta_3x_i)}) + e_i$	E - Meloun II	$y_i = \beta_1 - e^{(-\beta_2 - \beta_3x_i)} + e_i$
B - Gompertz	$y_i = \beta_1 e^{(-\beta_2 e^{(-\beta_3x_i)})} + e_i$	F - Brody	$y_i = \beta_1(1 - \beta_2 e^{(-\beta_3x_i)}) + e_i$
C - Logístico	$y_i = \frac{\beta_1}{(1 + \beta_2 e^{(-\beta_3x_i)})} + e_i$	G - von Bertalanffy	$y_i = \beta_1(1 - \beta_2 e^{(-\beta_3x_i)})^3 + e_i$
D - Meloun I	$y_i = \beta_1 - \beta_2 e^{(-\beta_3x_i)} + e_i$	H - Meloun III	$y_i = \beta_1 - (1 - \beta_3 e^{(-x_i - \beta_2)}) + e_i$

Os modelos foram ajustados através do método dos mínimos quadrados ordinários, utilizando o algoritmo iterativo de Gauss-Newton. Para todos os modelos, β_1 representa o peso assintótico; β_2 não apresenta interpretação prática, sendo apenas um parâmetro de locação; β_3 representa a velocidade de crescimento; e e_i é o erro aleatório, admitindo o pressuposto de que $e_i \sim N(0, \sigma^2)$, uma vez que valores discrepantes não foram detectados via análise gráfica.

2.2 Avaliadores de qualidade de ajuste

Para os modelos definidos, foram calculados cinco avaliadores de ajuste: o coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), desvio médio absoluto dos resíduos (DMA), e os critérios de informação de Akaike (AIC) e Bayesiano (BIC). Quanto maior o valor do coeficiente de determinação, melhor é o ajuste do modelo, enquanto que para os outros avaliadores menores valores representam melhores ajustes.

2.2.1 Coeficiente de determinação

Segundo Kennedy (2008), o coeficiente de determinação (R^2), deve representar a proporção da variação da variável resposta que é explicada pela variação da variável preditora. Existem várias dúvidas a respeito do uso do coeficiente de determinação para regressão não linear, sendo este sem significado para este caso de regressão, mas Souza (1998) afirma que a qualidade do ajustamento pode ser medida pelo quadrado do coeficiente de correlação entre os valores observados e preditos no caso de modelo de regressão não linear. O coeficiente de determinação pode ser calculado por:

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

onde SQR é a soma dos quadrados residuais; e SQT é a soma de quadrados total.

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

2.2.2 Quadrado médio do resíduo

O quadrado médio do resíduo (QMR) foi adotado como critério de verificação de qualidade de ajuste em estudos com regressão não linear em diversos trabalhos, como de Mazzini et al (2005) e Mello (2008). Este é dado por:

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

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

Menores valores do quadrado médio do resíduo indicam melhores ajustamentos.

2.2.1 Desvio médio absoluto dos resíduos

Proposto por Sarmiento (2006), o desvio médio absoluto dos resíduos (*DMA*), é calculado através do somatório dos desvios entre os pesos observados e estimados, dividido pelo número de observações, como indicado a seguir:

$$DMA = \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{n} \quad (3)$$

em que n é o número de observações; y_i é o valor observado da matéria seca total para a observação i ; e \hat{y}_i é o valor estimado da matéria seca total da observação i para o modelo em questão.

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

2.2.2 Critérios de informação de Akaike e Bayesiano

Os critérios de informação de Akaike (*AIC*) e Bayesiano (*BIC*), propostos por Akaike (1974) e Schwarz (1978), respectivamente, são estatísticas para comparação da qualidade de ajuste de modelo baseados no máximo da função de verossimilhança, sendo estes dependentes do número de observações e parâmetros do modelo em estudo. Assim, o valor do *AIC* é calculado por:

$$AIC = 2k - 2 \log L(\hat{\theta}) \quad (4)$$

e o valor de *BIC* é dado por:

$$BIC = k \log(n) - 2 \log L(\hat{\theta}) \quad (5)$$

nos quais n é o número de observações; k o número de parâmetros do modelo, e $L(\hat{\theta})$ o máximo da função de verossimilhança.

Segundo os critérios de informações de Akaike e Bayesiano, o melhor modelo é aquele que apresenta menor valor.

2.3 Agrupamento

Com os avaliadores calculados para cada modelo, foram então formados dados multivariados, nos quais cada modelo representa um indivíduo, e os avaliadores representam as variáveis. Para a análise de agrupamento, foi utilizado o método da

variância mínima de Ward (1963), com o número de grupos definido pelo método de Mojena (1977). A medida de dissimilaridade adotada foi a distância Euclidiana quadrática, sendo a distância Euclidiana originalmente a base para o método de Ward.

2.3.1 Método da variância mínima de Ward

Os modelos foram agrupados segundo o método da variância mínima de Ward (1963), método hierárquico que se baseia na mudança de variação dentro e entre grupos em formação a cada passo do processo. O procedimento é conhecido como método da variância mínima por ter como objetivo minimizar a soma de quadrados dentro dos grupos.

Para a formação inicial do grupo, este método considera o par de indivíduos que proporciona a menor soma de quadrados dos desvios. Em qualquer estágio, há perda de informações em razão do agrupamento realizado, podendo ser quantificada 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 do grupo é calculada considerando apenas os elementos dentro do grupo em formação, enquanto a soma de quadrados dos desvios total é calculada considerando todos os indivíduos disponíveis para o agrupamento. O agrupamento é realizado a partir do quadrado da distância Euclidiana, verificada a relação:

$$SQD_{i'v} = \frac{1}{2} d_{i'v}^2 \quad (6)$$

no qual

$$SQD_{i'v} = \sum_{j=1}^p SQD_{j(i'v)} \quad (7)$$

e

$$d_{i'v}^2 = \sum_{j=1}^p (Y_{ij} - Y_{i'v,j})^2 \quad (8)$$

em que $SQD_{j(i'v)}$ representa a soma de quadrados dos desvios para a j -ésima variável, considerando os modelos i e i' ; $d_{i'v}^2$ é o quadrado da distância Euclidiana entre os acessos i e i' ; p é o número de avaliadores de qualidade de ajuste utilizados; e Y_{ij} a média do caractere j para o modelo i .

A soma de quadrados dos desvios total, para g modelos avaliados, é dada por:

$$SQDT_{total} = \frac{1}{g} \sum_{i < i'}^g \sum_{i'}^g d_{i'v}^2 \quad (9)$$

Nesta análise, é identificado na matriz de quadrados das distâncias euclidianas $d_{i'v}^2$ (ou na matriz de somas dos quadrados dos desvios $SQD_{i'v}$) o par de modelos que resulta na menor soma de quadrados dos desvios. Agrupados estes modelos, uma nova matriz de dissimilaridade, de dimensão inferior, é recalculada, considerando que:

$$SQD_{(ijk)} = \frac{1}{k} d_{(ijk)}^2 \quad (10)$$

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

para k igual ao número de modelos no grupo, que neste caso é igual a três. E ainda:

$$SQD_{(ijkm)} = \frac{1}{k} d_{(ijkm)}^2 \quad (11)$$

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

agora para k igual a 4, e assim sucessivamente.

A família de algoritmos Lance e Williams (1967) ainda permite a definição e implementação do método de Ward. Esta família de algoritmos faz uso de uma fórmula de atualização da distância dos grupos em cada passo do agrupamento, tendo em vista a otimização da função objetivo do método. O algoritmo Lance e Williams para o método de Ward segue:

$$d_{(ij)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma |d_{ik} - d_{jk}| \quad (12)$$

no qual, para o algoritmo de Ward:

$$\alpha_i = \frac{n_i + n_k}{n_i + n_j + n_k}, \alpha_j = \frac{n_j + n_k}{n_i + n_j + n_k}, \beta = \frac{-n_k}{n_i + n_j + n_k} \text{ e } \gamma = 0 \quad (13)$$

onde d representa a distância entre as entidades recentemente agrupadas, i e j, e a entidade k; n_i , n_j e n_k representam o número de indivíduos em cada entidade.

Segundo Hair et al (2005), esse método tende a resultar em agrupamentos de tamanhos aproximadamente iguais devido a sua minimização de variação interna.

2.3.2 Critério de Mojena para determinação do número de grupos

O método de Ward gera dendrogramas, e o ponto de corte e número de grupos destes foram definidos pelo procedimento de Mojena (1977). Este método propõe um procedimento de cálculo baseado no tamanho relativo dos níveis de fusões (distâncias) do dendrograma, não necessitando do conhecimento prévio da conformação dos grupos (Faria, 2012). Assim, o procedimento visa selecionar o número de grupos no passo j que, primeiramente, satisfaça a seguinte inequação:

$$\alpha_j > \theta_k \quad (14)$$

na qual α_j é o valor de distância do nível de fusão correspondente aos passos j (para j=1, 2, ..., g-1), e θ_k é o valor referencial de corte, determinado por:

$$\theta_k = \bar{\alpha} + k \hat{\sigma}_\alpha \quad (15)$$

onde $\bar{\alpha}$ é a média de α ; $\hat{\sigma}_\alpha$ é o desvio padrão de α ; e k é uma constante, que define a regra de parada na definição do número de grupos, sendo adotado o valor de 1,25, conforme sugerido por Milligan e Cooper (1985). Dessa forma:

$$\bar{\alpha} = \frac{1}{g-1} \sum_{j=1}^{g-1} \alpha_j \quad (16)$$

e

$$\hat{\sigma}_\alpha = \sqrt{\frac{\sum_{j=1}^{g-1} \alpha_j^2 - \frac{1}{g-1} \left(\sum_{j=1}^{g-1} \alpha_j\right)^2}{g-2}} \quad (17)$$

O processo de agrupamento foi realizado para os dois grupos de modelos ajustados para os acessos formando, para cada um deles, *clusters* de modelos mais similares.

2.4 Identidade de modelos

Dentre os modelos ajustados, aquele com melhores resultados foi submetido a um teste de identidade de modelos (Regazzi; Silva, 2012), a fim de se avaliar se as estimativas dos parâmetros para os dois grupos de acessos são iguais. Foi realizado um teste F, no qual foi avaliada a hipótese de que o modelo reduzido ajustado para ambos os grupos é idêntico aos modelos completos ajustados. Assim:

$$F(H_0) = \frac{[SQRR(\omega) - SQRR(\Omega)]/[t(H-1)]}{SQRR(\Omega)/[N - Hp - H(r-1)]} \quad (18)$$

onde *SQRR* representa a soma de quadrados do resíduo da regressão (falta de ajustamento mais o resíduo) para dado modelo, Ω é espaço paramétrico para o modelo completo, ω é espaço paramétrico para o modelo reduzido sob H_0 , t é o número de parâmetros a serem testados, e N é o número total de observações. Para este teste foram utilizadas todas as observações ao invés de apenas as médias.

Todas as análises realizadas neste trabalho foram feitas utilizando o *software* R.

3 Resultados e discussão

Os modelos foram ajustados para os dois grupos (Tabela 3), sendo que apenas o modelo H (Meloun III) não convergiu para nenhum dos dois grupos de acessos. Os modelos A (Mitscherlich), D (Meloun I), E (Meloun II) e F (Brody) (Tabela 6), apresentam maiores valores para as estimativas do peso assintótico (β_1), enquanto os modelos B (Gompertz), C (Logístico) e G (von Bertalanffy) apresentam estimativas menores e mais próximas entre si, para ambos os grupos. As estimativas do parâmetro β_2 apresentam maiores variações entre os diferentes modelos, com os modelos C e D possuindo maiores valores. O modelo E, inclusive, estima um valor negativo para β_2 . Mas essa variação não tem importância, uma vez que este parâmetro tem apenas função de locação. As estimativas da taxa de crescimento (β_3) são viáveis para os modelos B, C e G, mas são muito pequenas para os demais modelos.

Tabela 3 - Estimativas dos parâmetros $\hat{\beta}_1$, $\hat{\beta}_2$ e $\hat{\beta}_3$ para os modelos dos dois grupos de acessos

Modelo	Grupo 1			Grupo 2		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
A – Mitscherlich	14738,3494	59,9764	0,0001	19088,3871	59,4267	0,0001
B – Gompertz	23,3109	475,4221	0,0625	21,9460	628,9370	0,0663
C – Logístico	22,8139	16357,5731	0,0915	21,6537	15849,0664	0,0926
D – Meloun I	19785,4437	19796,4650	0,0001	22272,7083	22283,8269	0,0001
E – Meloun II	3976,6158	-8,2938	0,0001	4510,4682	-8,4185	0,0001
F – Brody	15696,9776	1,0007	0,0001	18713,7078	1,0006	0,0001
G – von Bertalanffy	26,5807	10,5138	0,0359	24,9145	11,3016	0,0372

Os avaliadores de qualidade de ajuste são apresentados nas Tabelas 4 e 5. O grupo 2 apresentou melhores resultados quanto ao valor da média, e também menores coeficientes de variação. O coeficiente de determinação (R^2) foi o avaliador com menor variação dentre os modelos. Essa baixa variabilidade do R^2 quando aplicado a regressão não linear é comum, e segundo Oliveira et al (2000), as diferenças neste coeficiente entre modelos usados no estudo de curvas de crescimento são mínimas, fazendo-se necessária a utilização de outros avaliadores. Os valores de quadrado médio do resíduo (QMR) e desvio médio absoluto dos resíduos (DMA) apresentaram as maiores variações; isso pode ser visto nos dois grupos de acessos que possuem valores bem extremos, embora estes valores sejam próximos se comparando um mesmo modelo entre os dois grupos. As informações de AIC e BIC penalizam modelos com muitas variáveis, mas todos os modelos ajustados possuem mesmo número de parâmetros e variáveis.

Tabela 4 - Estimativas dos avaliadores de qualidade de ajuste para os modelos, e respectivas média, desvio padrão e coeficiente de variação (CV%) do grupo 1 de acessos

Modelo	Avaliadores de qualidade de ajuste				
	R^2	QMR	DMA	AIC	BIC
A – Mitscherlich	0,9325	22,0244	1,9789	26,1749	23,7201
B – Gompertz	0,9977	0,7511	0,2173	12,6617	10,2068
C – Logístico	0,9991	0,2987	0,1793	8,9731	6,5182
D – Meloun I	0,9104	29,2282	2,3346	27,3069	24,8521
E – Meloun II	0,8956	34,0562	2,6913	27,9184	25,4636
F – Brody	0,8769	40,1289	2,8061	28,5747	26,1199
G – von Bertalanffy	0,9875	4,0676	0,9750	19,4185	16,9637
Média	0,9428	18,6507	1,5975	21,5755	19,1206
Desvio padrão	0,0515	16,7917	1,1292	8,0224	8,0224
CV%	5,4616	90,0323	70,6828	37,1831	41,9569

Tabela 5 - Estimativas dos avaliadores de qualidade de ajuste para os modelos, e respectivas média, desvio padrão e coeficiente de variação (CV%) do grupo 2 de acessos

Modelo	Avaliadores de qualidade de ajuste				
	R^2	QMR	DMA	AIC	BIC
A – Mitscherlich	0,9299	20,4701	1,8233	25,8822	23,4274
B – Gompertz	0,9962	1,1211	0,2651	14,2637	11,8089
C – Logístico	0,9982	0,5344	0,2456	11,3002	8,8454
D – Meloun I	0,9175	24,1165	1,9761	26,5379	24,0831
E – Meloun II	0,8931	31,2219	2,4483	27,5708	25,1160
F – Brody	0,8845	33,7443	2,4485	27,8816	25,4268
G – von Bertalanffy	0,9842	4,6270	1,0407	19,9340	17,4791
Média	0,9434	16,5479	1,4639	21,9101	19,4552
Desvio padrão	0,0488	14,2609	0,9513	6,8293	6,8293
CV%	5,1736	86,1793	64,9847	31,1699	35,1028

Para os dois grupos avaliados, os modelos foram particionados, seguindo o critério proposto por Mojena (1970), em dois grupos (Figura 1). As partições obtidas foram as mesmas para os dois grupos, sendo o *cluster 1* formado pelos modelos B (Gompertz), C (Logístico) e G (von Bertalanffy), e o *cluster 2* pelos modelos A (Mitscherlich), D (Meloun I), E (Meloun II) e F (Brody) (Tabela 6). O *cluster 1* apresentou melhores resultados quanto aos avaliadores de ajustamento (maior valor de R^2 , menores valores de *QMR*, *DMA*, *AIC* e *BIC*), além de estimativas de parâmetros viáveis para interpretação prática. Já os parâmetros dos modelos do *cluster 2* tiveram grande variação entre si e pouco sentido para interpretação biológica, como por exemplo o modelo A (Mitscherlich) do grupo 1 que apresenta valor muito grande de β_1 ($\hat{\beta}_1 = 14738,3494$), e valor muito pequeno de β_3 ($\hat{\beta}_3 = 0,0001$).

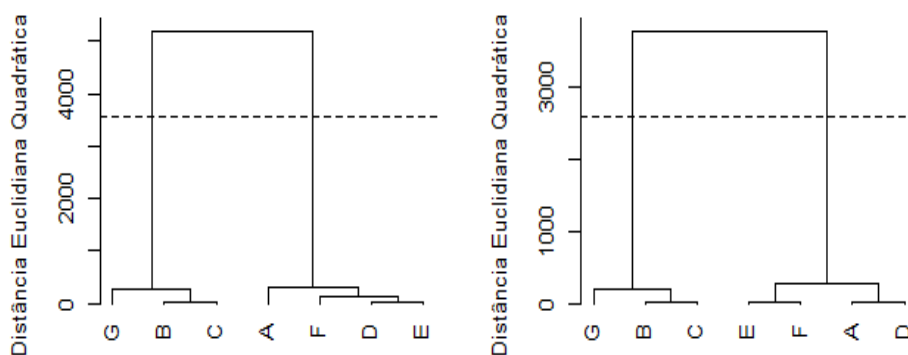


Figura 1 - Dendrogramas para o agrupamento dos modelos do grupo 1 (esquerda) e grupo 2 (direita), obtidos com o algoritmo de Ward.

Tabela 6 - *Clusters* formados, com seus modelos e médias de avaliadores de qualidade de ajuste, para os dois grupos de acessos

Grupo	Avaliadores	Cluster 1	Cluster 2
		Modelos: B, C, G.	Modelos: A, D, E, F.
1	R ²	0,9948	0,9038
	QMR	1,7058	31,3594
	DMA	0,4572	2,4527
	AIC	13,6844	27,4937
	BIC	11,2296	25,0389
2	R ²	0,9928	0,9063
	QMR	2,0942	27,3882
	DMA	0,5172	2,1740
	AIC	15,1660	26,9681
	BIC	12,7111	24,5133

Observando os modelos pertencentes ao *cluster* 1, o modelo C (Logístico) é o melhor para a descrição do acúmulo de matéria seca para os dois grupos de acessos de alho. Este modelo, além de melhor em termos dos resultados dos avaliadores de qualidade de ajuste, é um dos mais utilizados para descrição de crescimento em plantas, inclusive de alho (Reis, 2012). Logo, o modelo C foi selecionado para realização dos testes de identidade. As curvas do modelo Logístico para os dois grupos de modelos em relação à matéria seca total da planta, são apresentadas na Figura 2.

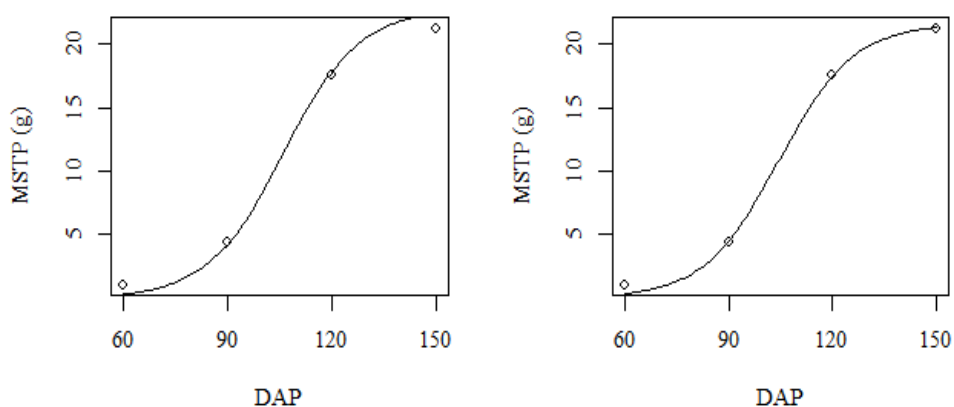


Figura 2 - Curvas do acúmulo de matéria seca total da planta (MSTP) ajustadas pelo modelo Logístico no grupo 1 (esquerda) e no grupo 2 (direita).

Os testes de identidade de modelos foram realizados sobre os parâmetros com interpretação biológica: o peso assintótico (β_1) e a velocidade de crescimento (β_3). São apresentadas na Tabela 7 as estimativas dos parâmetros do modelo sem restrições no espaço paramétrico (modelo completo Ω), e dos modelos reduzidos segundo as restrições ω_1 , ω_2 e ω_3 .

Tabela 7 - Estimativas dos parâmetros do modelo completo (Ω) e dos modelos reduzidos (ω_1 , ω_2 e ω_3)

Parâmetro	Modelo			
	Ω	ω_1	ω_2	ω_3
$\hat{\beta}_1$	-	-	22,2388	22,2330
$\hat{\beta}_{1(1)}$	22,8199	22,6294	-	-
$\hat{\beta}_{1(2)}$	21,6479	21,8431	-	-
$\hat{\beta}_2$	16112,0883	15962,0112	15868,7516	15955,9977
$\hat{\beta}_3$	-	0,0920	-	0,0920
$\hat{\beta}_{3(1)}$	0,0914	-	0,0921	-
$\hat{\beta}_{3(2)}$	0,0928	-	0,0917	-

Foram testadas a hipóteses de que os modelos reduzidos ajustados para os dois grupos de acessos são idênticos ao modelo completo. O experimento possui dois grupos de 10 acessos cada, em quatro blocos, com duas plantas do acesso por bloco, e avaliados em quatro períodos, em um total de 640 observações. De acordo com os resultados (Tabela 8), não houve diferença entre as estimativas para o peso assintótico do grupo 1 ($\hat{\beta}_{1(1)} = 22,8139$) para o grupo 2 ($\hat{\beta}_{1(2)} = 21,6537$), nem para a velocidade de crescimento (onde $\hat{\beta}_{3(1)} = 0,0915$ e $\hat{\beta}_{3(2)} = 0,0926$). Assim, uma única equação completa pode representar os dois grupos de acessos.

Tabela 8 - Testes de identidade de modelos pela estatística F. Os parâmetros testados foram aqueles com interpretação biológica (β_1 e β_3) do modelo Logístico, para $\alpha = 0,05$

Hipóteses	G. L.	F _{calc}	p-valor
$H_0^{(1)}: \beta_{1(1)} = \beta_{1(2)} = \beta_1$ vs $H_A^{(1)}$	1 e 635	1,4620	0,2271 ns
$H_0^{(2)}: \beta_{3(1)} = \beta_{3(2)} = \beta_3$ vs $H_A^{(2)}$	1 e 635	0,3635	0,5468 ns
$H_0^{(3)}: \beta_{1(1)} = \beta_{1(2)} = \beta_1$ e $\beta_{3(1)} = \beta_{3(2)} = \beta_3$ vs $H_A^{(3)}$	2 e 635	0,7652	0,4657 ns

O teste de $H_0^{(3)}$ foi não significativo ($P > 0,05$). Logo, a equação comum, apresentada pelo modelo ω_3 (Tabela 7), pode ser adotada para os dois grupos de acessos. O modelo escolhido é o com três parâmetros β_1 , β_2 e β_3 , sendo a equação ajustada:

$$\hat{y} = \frac{22,233}{(1 + 15955,9977e^{(-0,092x)})} \quad (19)$$

Conclusões

A análise de agrupamento forneceu, para ambos os grupos de acessos, dois *clusters* de modelos mais próximos. Para os dois grupos o primeiro *cluster*, formado pelos modelos de Gompertz, Logístico e von Bertalanffy, mostrou melhores resultados para descrever o acúmulo de matéria seca total da planta para os acessos de alho estudados.

Dentre estes, o modelo Logístico se destacou segundo os avaliadores de qualidade de ajuste, e a técnica de identidade de modelos mostrou que foi possível o ajuste de uma equação única para os dois grupos.

Agradecimentos

Os autores agradecem à FAPEMIG e ao CNPq pelos recursos para realização deste trabalho.

PUIATTI, G. A.; CECOM, P. R.; NASCIMENTO, M.; PUIATTI, M.; FINGER, F. L.; SILVA, A. R. da; NASCIMENTO, A. C. C. Cluster analysis applied to nonlinear regression models selection for the description of dry matter accumulation of garlic plants. *Rev. Bras. Biom.*, São Paulo, v.31, n.3, p.337-351, 2013.

- **ABSTRACT:** *This work had as objective the evaluation of nonlinear regression used for description of the dry matter accumulation along the time in plants of garlic, comparing results of quality fit evaluators. Twenty garlic accessions were divided in two groups, evaluated in the characteristics: diameter, length and medium weight of the bulb, number of bulbils and productivity. For the two groups, nonlinear regression models were adjusted to describe the dry matter accumulation of the plant. The models were then grouped according to the fit quality, measured by the coefficient of determination, error mean square, average deviation absolute error, Akaike information criterion, and Bayesian information criterion. The analysis indicated the Logistic model as the most appropriate for the description of dry matter accumulation in garlic, and a test of model identity applied to the models of the two groups indicated that they are statistically similar, and a complete model is valid for both groups of garlic accessions.*
- **KEYWORDS:** *Multivariate classification; model identity; Allium sativum L.*

Referências

- 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, 1974.
- FARIA, P. N.; CECOM, P. R.; SILVA, A. R.; FINGER, F. L.; SILVA, F. F.; CRUZ, C. D.; SÁVIO, F. L. Métodos de agrupamento em estudo de divergência genética de pimentas. *Hortic. Bras.*, Vitória da Conquista, v. 30, n. 3, p. 428-432, 2012.
- HAIR, J. F.; et al. *Análise multivariada de dados*. Trad. Adonai S. Sant'Anna e Anselmo C. Neto. 5 ed. Porto Alegre: Bookman, 2005.

- HONORATO, A. R. de F. *Avaliação de cultivares de alho na região de Mossoró-RN*. 2012. 59f. Dissertação (Mestrado em Fitotecnia) – Universidade Federal Rural do Semi-Árido, Mossoró, 2012.
- JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. New Jersey, USA: Englewood Cliffs, 1992. 642p.
- KENNEDY, Peter. *A Guide to Econometrics*. Massachusetts: Blackwell Publishing, 2008.
- LANCE, G. N.; WILLIAMS, W. T. A General Theory of Classificatory Sorting Strategies. 1. Hierarchical Systems, *Computer Journal*, v.9, p.373-380, 1967.
- 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.
- MAZZINI, A. R. A.; MUNIZ, J. A.; SILVA, F. F. e; AQUIN, L. H. Curvas de crescimento de novilhos Hereford: heterocedasticidade e resíduos autorregressivos. *Ciência Rural*, Santa Maria, v. 35, n. 2, p. 422-427, 2005.
- MELLO, R.; MAGALHAES, A. L. R.; BREDA, F. C.; REGAZZI, A. J. Modelos para ajuste da produção de gases em silagens de girassol e milho. *Pesq. agropec. bras.*, Brasília, v. 43, n. 2. 2008.
- 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.
- MOJENA, R. Hierárquical grouping method and stopping rules: an evaluation. *Computer Journal*, v.20, p.359-363, 1977.
- OLIVEIRA, H. N. de; LOBÔ, R. B.; PEREIRA, C. S. Comparação de modelos não-lineares para descrever o crescimento de fêmeas da raça Guzerá. *Pesquisa Agropecuária Brasileira*, v.35, n.9, p.1843-1851, 2000.
- R DEVELOPMENT CORE TEAM. R: *A language and environment for statistical computing*. Viena: R Foundation for Statistical Computing, 2004.
- REGAZZI, A. J.; SILVA, C. H. O. Testes para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear em dados de experimento com delineamento em blocos casualizados. *Revista Ceres*, Viçosa, v. 57, n. 3, p. 315-320, 2010.
- REIS, R. M. dos. *Modelos de regressão não linear para descrição do crescimento de plantas de alho*. 2012. 47f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa, 2012.
- 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, 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.

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

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.

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

Recebido em 10.09.2013

Aprovado após revisão em 11.11.2013