



WALMES MARQUES ZEVIANI

**PARAMETRIZAÇÕES INTERPRETÁVEIS
EM MODELOS NÃO LINEARES**

**LAVRAS - MG
2013**

WALMES MARQUES ZEVIANI

**PARAMETRIZAÇÕES INTERPRETÁVEIS EM MODELOS NÃO
LINEARES**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientador
Prof. Dr. Joel Augusto Muniz

**LAVRAS - MG
2013**

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Zeviani, Walmes Marques.

Parametrizações interpretáveis em modelos não lineares / Walmes
Marques Zeviani. – Lavras : UFLA, 2013.

146 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2013.

Orientador: Joel Augusto Muniz.

Bibliografia.

1. Verossimilhança. 2. Método delta. 3. Medida de curvatura. 4.
Efeitos mistos. 5. van Genuchten. I. Universidade Federal de Lavras.
II. Título.

CDD-519.76

WALMES MARQUES ZEVIANI

**PARAMETRIZAÇÕES INTERPRETÁVEIS EM MODELOS NÃO
LINEARES**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 13 de maio de 2013.

Prof. Dr. Paulo Justiniano Ribeiro Jr	UFPR
Prof. Dr. Júlio da Motta Singer	IME-USP
Prof. Dr. Júlio Silvio de Sousa Bueno Filho	UFLA
Prof. Dr. Augusto Ramalho de Moraes	UFLA

Prof. Dr. Joel Augusto Muniz
Orientador

**LAVRAS - MG
2013**

À DEUS, autor da minha vida.

*Aos meus pais, Jaime e Marilene, por todo amor, cuidado,
apoio, respeito, confiança e paciência;*

Aos meus irmãos, Wolnei e Waires, pela admiração, respeito e amizade.

DEDICO

AGRADECIMENTOS

À meu pai, Jaime, por toda confiança e incentivo que permitiram alcançar esse objetivo e a minha mãe, Marilene, por todo carinho, apoio e infância que me proporcionou.

Aos meus irmãos, Wolnei e Waires, pelo apoio, admiração, respeito e incentivo.

Aos meus avós, tios e primos, pelo carinho e suporte.

Ao professor Joel Augusto Muniz, pelos conselhos, paciência e dedicação durante todo o desenvolvimento da orientação.

À professora Yara, pelas orientações durante a graduação em Agronomia e direcionamento da minha carreira acadêmica.

Ao professor João Dimas, enquanto tutor do PET Agronomia, pela oportunidade e incentivo à pesquisa.

Ao amigo e Professor Wagner Hugo Bonat pela fundamental colaboração, incentivo, amizade e momentos de descontração proporcionados.

Ao Professor Paulo Justiniano por todo ensinamento, palavras de sabedoria e incentivo que permitiram a conclusão dessa Tese em paralelo ao meu aperfeiçoamento pessoal e profissional.

Aos demais Professores do LEG, Silvia Shimakura e Elias Krainski, pelas sugestões durante seminários sobre o Tema.

Aos colegas de profissão do DCET da UESC, Enio Jelihovschi, Luiz Roberto, JC Faria, Ivan Allaman e Fábio Corrêa, pelo entusiasmo sobre a primeira exposi-

ção do Tema dessa Tese em forma de curso e incentivo.

Aos Professores Júlio da Motta Singer, Júlio Silvio de Sousa Bueno Filho e Geraldo Cesar de Oliveira, pelas críticas e sugestões durante Exame de Qualificação.

À Dania Messmar, Carla Carducci, Rosiana Rodrigues e Manoel Victor pela sincera amizade.

Aos companheiros de república em Curitiba, Maciel Batista e Cláudio Afonso, pelo excelente convívio.

Aos companheiros da república Mete-Caixa durante moradia em Lavras, em especial aos amigos Fábio Ono e Glauco Teixeira, pelo convívio e acolhimento nas minhas visitas.

Aos amigos Milson Serafim e Izidro Lima Jr, pela colaboração científica e concessão dos dados.

À todos os meus colegas de mestrado e doutorado por compartilharem momentos de aprendizado e descontração.

À minha namorada, pela admiração, respeito e amor.

À turma dos treinos de Muay Thai, pelas lições de disciplina, coragem e resistência.

Aos meus alunos pelo meu aperfeiçoamento como Professor.

Aos participantes dos meus cursos de capacitação, pela troca de conhecimento e colaborações de pesquisa.

À lista de discussão brasileira de R, em especial aos membros Benilton Carvalho e Henrique Dallazuanna pelas sugestões de implementação.

Aos desenvolvedores de software livre, por fornecerem as ferramentas necessárias para o desenvolvimento dessa Tese.

À UFPR pela oportunidade de exercer minha vocação de transmitir conhecimento e produzir pesquisa científica.

À CAPES pela bolsa de Mestrado e pela bolsa de mudança de nível para o Doutorado.

“Experiência não é o que acontece com você, mas o que você fez com o que lhe aconteceu.”

Aldous Huxley

“O que você sabe não tem valor, o valor está no que você faz com o que sabe.”

Bruce Lee

“Quem não sabe o que busca não identifica o que acha.”

Immanuel Kant

RESUMO GERAL

Uma das vantagens dos modelos de regressão não linear é ter interpretação para os parâmetros. Em muitas situações, parâmetros de interesse, expressos como função dos parâmetros do modelo, são quantidades sujeitas à investigação. Surge então a preocupação de como fazer inferência sobre eles. Para isso, o método delta, a simulação Monte Carlo e procedimentos bootstrap são alternativas frequentes. Além disso, uma reparametrização pode ser aplicada ao modelo de forma à representar tais parâmetros de interesse. Além de melhorar a interpretação, a presença do parâmetro alvo estende as possibilidades com relação a especificação de modelos e inferência estatística. O objetivo com esse trabalho é sistematizar o procedimento de aplicar reparametrizações. Ênfase foi dada em modelos não lineares considerados em aplicações dentro das Ciências Agrárias. Uma lista com 17 modelos reparametrizados é fornecida. No primeiro estudo de caso, o nível de dano econômico da desfolha no algodoeiro foi avaliado com os seguintes objetivos: 1) propor uma parametrização de modelo que representasse o nível de dano econômico, 2) avaliar parametrizações alternativas por meio de suas propriedades, onde considerando medidas de não linearidade, 3) aplicar inferência baseada em verossimilhança, 4) selecionar um modelo para descrever a relação entre produção e desfolha do algodoeiro em função do estágio fenológico. O modelo reparametrizado apresentou melhores propriedades nos estágios fenológicos com pronunciada relação não linear. No restante, as medidas de curvatura, as correlações dos estimadores e os gráficos de perfil de verossimilhança indicaram que um sub-modelo deveria ser considerado. No segundo estudo de caso, objetiva-se verificar o efeito da posição de amostragem e profundidade do solo sobre os parâmetros I (*inflection*) e S (*slope*) da curva de retenção de água do solo. Para isso 1) considerou-se ANOVA simples e 2) ANOVA ponderada pela variância das estimativas desses parâmetros em cada unidade experimental em comparação com 3) o uso de modelos não lineares de efeitos mistos em uma parametrização desenvolvida. Nenhum dos métodos alternativos de análise foi superior ao modelo não linear de efeitos mistos na parametrização desenvolvida, que apresentou intervalos de confiança mais estreitos para os parâmetros e apontou efeito de posição e profundidade de coleta.

Palavras-chave: Verossimilhança. Método delta. Medidas de curvatura. Efeitos mistos. van Genuchten.

GENERAL ABSTRACT

One of the advantages of the nonlinear regression models is to have interpretable parameters. In many instances, the parameters of interest, expressed as a function of the model parameters, are quantities subject to investigation. Then comes the concern of how to make inferences about them. For this, the delta method, the Monte Carlo simulation and bootstrap procedures are common alternatives. In addition, a reparametrization can be applied to the model in order to represent these parameters of interest into the model. In addition to improving the interpretation of the presence of the target parameter extends the possibilities regarding the specification of models and statistical inference. The aim of this work is to systematize the procedure to apply reparametrizations. Emphasis was given on nonlinear models considered in applications within the Agricultural Sciences. A list with 17 models reparametrized is provided. In the first case study, the threshold level of defoliation on cotton was evaluated with the following objectives: 1) to propose a model parameter that represents the level of economic damage, 2) evaluate alternative parameterizations through its properties, which considering measures of nonlinearity, 3) apply inference based on likelihood, 4) select a model to describe the relationship between yield and defoliation of cotton in each phenological stage. The reparametrized model showed better properties in phenological stages with pronounced nonlinear relationship. Otherwise the measures of curvature, the correlations of the estimators and likelihood profile plots indicated that a sub-model should be considered. In the second case study, the objective is to verify the effect of sampling position and soil depth on the parameters I (*inflection*) and S (*slope*) of the soil water retention curve. For that 1) it was considered ANOVA and 2) weighted ANOVA in each experimental unit compared to 3) using nonlinear mixed effects on a parameterization developed. None of the alternative methods of analysis was superior to model nonlinear mixed effects in the parameterization developed, which had narrower confidence intervals for the parameters and pointed sampling position and depth effect.

Keywords: Likelihood. Delta method. Curvature measures. Mixed effects. van Genuchten.

SUMÁRIO

CAPÍTULO 1 Procedimento para obter parametrizações interpretáveis em modelos não lineares		14
1	INTRODUÇÃO	16
2	REPARAMETRIZAÇÃO	20
2.1	Reparametrização 1:1 - Modelos para acúmulo com ênfase na fração do total.	22
2.2	Reparametrização 2:2 - Modelos com ênfase no ponto crítico.	25
2.3	Excesso de parâmetros de interesse.	28
2.4	Uso de soluções numéricas	31
2.5	Ponto crítico em polinômios de grau 2 e 3	32
2.6	Distribuições de probabilidade	34
2.7	Outros modelos	36
3	ESTIMAÇÃO.	48
3.1	Verossimilhança	48
3.2	Aproximação quadrática da verossimilhança	50
3.3	Método delta	51
4	CONSIDERAÇÕES FINAIS	54
	REFERÊNCIAS.	57
CAPÍTULO 2 Modelo não linear para o nível de dano econômico da desfolha no algodoeiro		63

1	INTRODUÇÃO	65
2	MODELO	68
3	MATERIAL E MÉTODOS.	72
4	RESULTADOS E DISCUSSÃO	74
5	CONCLUSÕES	85
	REFERÊNCIAS.	86
	ANEXOS	90
	CAPÍTULO 3 Parametrização do modelo van Genuchten para inferência sobre os parâmetros S e I	95
1	INTRODUÇÃO	97
2	RETENÇÃO DE ÁGUA E TAMANHO DE POROS	105
3	REPARAMETRIZAÇÃO	109
4	MATERIAL E MÉTODOS.	113
4.1	Medidas de curvatura.	114
4.2	Análises em dois estágios	117
4.3	Modelo não linear de efeitos mistos (NLMEM)	117
5	RESULTADOS E DISCUSSÃO	121
6	CONCLUSÕES	136
	REFERÊNCIAS.	137
	ANEXOS	141

CAPÍTULO 1 PROCEDIMENTO PARA OBTER PARAMETRIZAÇÕES INTERPRETÁVEIS EM MODELOS NÃO LINEARES

RESUMO

Modelos de regressão não linear são considerados quando existe algum conhecimento preliminar sobre a relação entre variáveis. Tal conhecimento pode ser a respeito da própria natureza dos dados, uma equação diferencial e até mesmo a forma do diagrama de dispersão entre as variáveis. Em geral, seus parâmetros têm interpretação. Além disso, parâmetros de interesse, expressos como função dos parâmetros do modelo, são alvos de investigação. Para isso, o método delta, simulação Monte Carlo e procedimentos bootstrap são procedimentos adotados para fazer inferência. Além disso, uma reparametrização pode ser aplicada ao modelo de forma a representar esses parâmetros de interesse. Além de melhorar a interpretação do modelo, a presença do parâmetro alvo estende as possibilidades com relação a especificação de modelos e inferência estatística. O objetivo com esse trabalho é sistematizar o procedimento de aplicar reparametrizações. Ênfase é dada em modelos não lineares considerados em Ciências Agrárias. Uma lista com 17 modelos reparametrizados é fornecida. Breve discussão sobre os métodos de inferência é feita.

Palavras-chave: Função de parâmetros. Interpretação de parâmetros. Verossimilhança. Método delta. Curvatura.

ABSTRACT

Nonlinear regression models are considered when there is some prior knowledge of the relationship between variables. Such knowledge can be about the nature of the data, a differential equation, and even the shape of the scatter diagram between the variables. In general, the parameters have an interpretation. Furthermore, parameters of interest, expressed as a function of the model parameters, are targets of investigation. For this, the delta method, Monte Carlo simulation and bootstrap procedures are procedures used to make inferences. In addition, a reparametrization can be applied to the model to represent the parameters of interest. In addition to improving the interpretation of the model, the presence of the target parameter extends the possibilities regarding the specification of models and statistical inference. The aim of this work is to systematize the procedure to apply reparametrizações. Emphasis is on nonlinear models considered in Agricultural Sciences. A list with 17 models reparametrized is provided. Brief discussion on the methods of inference is made.

Key-words: Function of parameters. Parameter interpretation. Likelihood. Delta Method. Curvature.

1 INTRODUÇÃO

A ideia básica da regressão não linear é a mesma da regressão linear: relacionar uma resposta Y com um vetor de variáveis preditoras $\mathbf{x} = (x_1, \dots, x_k)^\top$. Os modelos de regressão não linear são caracterizados pelo fato de a função de predição depender não linearmente de algum dos parâmetros. Embora não necessariamente, a regressão linear é usada para especificação de modelos puramente empíricos, enquanto que os modelos de regressão não linear são considerados quando existe algum conhecimento prévio para sustentar que a relação entre resposta e preditores segue uma particular forma funcional. Tal conhecimento pode ser desde uma equação diferencial que remete à particular modelo, como é o caso de modelos de crescimento, ou simplesmente uma restrição sobre a função, como o de a função ser monótona, típico de curvas de acúmulo, para a qual pode-se ter várias funções disponíveis.

Uma das principais vantagens do modelo de regressão não linear, é que frequentemente existe interpretação para a maioria de seus parâmetros (SCHABENBERGER; PIERCE, 2002). Esses parâmetros então passam ser o foco da investigação que, na sua forma mais simples, consiste em determinar intervalos de confiança e testar hipóteses. No entanto, uma situação comum é a necessidade de fazer inferência sobre uma função dos parâmetros (BENDER, 1996). Um exemplo simples é a equação de segundo grau $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$, que é um modelo linear no qual o ponto crítico $x_c = -\beta_1/(2\beta_2)$ é alvo de inferência em situações de otimização de processos (BAS; BOYACI, 2007). Uma vez estimados seus parâmetros, inferência sobre x_c pode ser feita pelo método delta, por simulação Monte Carlo ou por métodos *bootstrap* (SEBER; WILD, 2003). Embora tais procedimentos permitam obter intervalos de confiança e conduzir testes de hipótese, existem

ainda outras formas vantajosas de inferir ou modelar o parâmetro que são as extensões ligadas aos modelos não lineares de efeitos mistos (PINHEIRO; BATES, 2000) e a inferência bayesiana (DENISON et al., 2002).

Nos modelos não lineares de efeitos mistos, a especificação completa do modelo se dá em pelo menos dois níveis. No primeiro, especifica-se um modelo não linear $f(x, \theta)$ para a média de Y . No segundo nível, especifica-se um modelo para θ , que pode considerar fatores de efeito fixo e aleatório. Então que pode-se medir diretamente o efeito de um fator sobre Y por meio do seu efeito sobre alguns parâmetros em θ . No escopo da inferência bayesiana é possível incorporar o conhecimento do pesquisador sobre os parâmetros por meio de prioris. Com isso, o ganho em termos de capacidade de especificação do modelo, permitem que o pesquisador represente melhor o processo de geração dos dados, seja pela adequada descrição do delineamento experimental, pela incorporação de conhecimento a priori ou por considerar covariáveis e fatores. Entretanto, para ter todos esses benefícios ligados à uma particular interpretação o parâmetro de interesse deve ser componente do vetor de parâmetros do modelo.

A reparametrização permite introduzir p parâmetros de interesse ao custo da substituição de p dos atuais parâmetros. As reparametrizações em modelos de regressão não linear são em sua maioria adotadas para mapear espaços paramétricos limitados à reta real facilitando os procedimentos de otimização (RATKOWSKY, 1990; CHENG et al., 1992), para impor restrições paramétricas, para diminuir problemas de identificabilidade (ROSS; SARADA, 2010; MESHKAT et al., 2011), para amenizar os efeitos da não linearidade do modelo (BATES; WATTS, 1980; HOUGAARD, 1982; RATKOWSKY, 1986; HOUGAARD, 1988; TSAI, 1988; BARROZO et al., 2004; SARADA, 2005; ROSS; SARADA, 2010), para obter delineamentos ótimos (FORD et al., 1989; PÁZMAN; PRONZATO,

1992; ATKINSON et al., 1993; VILA; GAUCHI, 2007), e inclusive, como é o nosso objetivo, para melhorar a interpretação nos parâmetros.

Segundo Ross e Sarada (2010), uma reparametrização pode ser considerada tanto para fornecer uma descrição funcional simples, quanto para fornecer estimativas diretas de parâmetros com significado provenientes do processo biológico subjacente. Por meio de reparametrização Rekaya et al. (2001) incorporaram o parâmetro persistência de lactação ($P = C^{-B+1}$) ao modelo de lactação de Wood (WOOD, 1967) ($f(x) = Ax^B \exp\{-Cx\}$) e com isso foram capazes de usar de informações de *pedigree* ao estimar de P , além de acomodar o efeito das diferenças entre animais e devido ao fator ordem de lactação por meio de um modelo não linear bayesiano hierárquico. Dhanoa (1981) por sua vez, reparametrizou este modelo de lactação ($Ax^{mC} \exp\{-Cx\}$) para contemplar o tempo necessário para atingir a produção máxima (m) e ainda observou redução das correlações entre os parâmetros do modelo, o é vantajoso por acelera procedimentos de estimação. Lyles et al. (2008) reparametrizaram o modelo exponencial, logístico e Gompertz para estimar o parâmetro IC_{50} , que representa a concentração de um composto para qual se tem uma inibição de 50% da resposta. Zeviani et al. (2012), estenderam e aplicaram reparametrização ao modelo exponencial assintótico (de $f(x) = A(1 - \exp\{-Bx\})$ para $A(1 - \exp\{-\log(2)x/V\})$) para ter interpretação idêntica ao modelo Michaelis-Menten ($f(x) = Ax/(V + x)$), em que A é o conteúdo total de nutriente a liberar para o solo e V é o tempo para liberação de 50% de A , e assim selecionar o modelo com melhores propriedades, do ponto de vista de inferência, para representar a liberação de potássio de esterco de animais incorporados em latossolos.

Embora exista considerável literatura relacionada à reparametrização de modelos de regressão não linear, ainda é restrita aquela direcionada à melhoria na interpretação do modelo. Dados os benefícios da reparametrização com essa

finalidade, relacionados à inferência baseada em verossimilhança, à capacidade de especificação via modelos não lineares de efeitos mistos, à atribuição de priors e o planejamento de delineamentos ótimos, o objetivo com este trabalho é 1) sistematizar o procedimento de tal forma que possam ser reproduzido em diversos contextos 2) exemplificar por meio de aplicações as vantagens, cuidados e desdobramentos da reparametrização visando interpretação, para que se possa ter direcionamentos com relação à sua adoção, 3) comparar as inferências obtidas sob a reparametrização com àquelas frequentemente utilizadas para fazer inferência sobre uma função de parâmetros.

Na próxima seção são apresentados os passos para aplicar reparametrização em um modelo. Uma série de modelos foi considerada para a demonstração do procedimento.

2 REPARAMETRIZAÇÃO

Considere um modelo não linear

$$f(\mathbf{x}, \boldsymbol{\theta}) \tag{2.1}$$

em que f é uma função não linear que depende do vetor de covariáveis \mathbf{x} e $\boldsymbol{\theta}$ é seu vetor de p parâmetros. Seja $\vartheta = g(\boldsymbol{\theta})$ o parâmetro de interesse em que g é uma função monótona e diferenciável em relação à $\boldsymbol{\theta}$. O objetivo com a reparametrização é fazer com que ϑ seja um elemento do vetor de parâmetros do modelo. Isso é obtido por substituição de algum dos p elementos de $\boldsymbol{\theta}$ por ϑ . Para isso, sistematizou-se o procedimento em três etapas:

1. Expressar o parâmetro de interesse como função dos elementos de $\boldsymbol{\theta}$, ou seja, $\vartheta = g(\boldsymbol{\theta})$;
2. Escolher um dos elementos θ_i de $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}_{-i})$ para ser colocado em função de ϑ de tal forma a obter $\theta_i = h(\boldsymbol{\theta}_{-i}, \vartheta)$;
3. Substituir θ_i em (2.1) pela expressão obtida no passo anterior, $h(\boldsymbol{\theta}_{-i}, \vartheta)$, fazendo as simplificações convenientes. Assim o modelo (2.1) pode ser expresso como

$$f(\mathbf{x}, \boldsymbol{\theta}_{-i}, \vartheta)$$

A função h é a inversa de g em θ_i .

No passo 2 recomenda-se priorizar aquele elemento de $\boldsymbol{\theta}$ com menor significado ou para o qual a obtenção de h é menos complicada. Em algumas situações pode ser que h não tenha solução analítica, por exemplo quando é impossível isolar

θ_i , e assim uma solução numérica pode ser usada para obter o modelo reparametrizado. O procedimento também pode ser usado para mais de um parâmetro de interesse, basta repeti-lo até que todos $(\vartheta_1, \dots, \vartheta_k, k \leq p)$ sejam incorporados. Schabenberger et al. (1999) consideraram essas etapas no estudo de dose-resposta com herbicidas sob o modelo log-logístico, o qual foi reparametrizado para incluir o parâmetro λ_k , que representa a dose corresponde à uma fração $0 < k < 1$ da diferença entre a assíntota superior e inferior. Da mesma forma, Schabenberger e Pierce (2002) apresentam uma versão reparametrizada do modelo Mitscherlich que contém o índice de disponibilidade do nutriente, útil no contexto de nutrição de plantas.

Uma etapa importante na reparametrização é fazer o estudo de dimensionalidade para determinar as unidades de medida dos parâmetros bem como seus espaços paramétricos. Tais informações são relevantes para descrição do modelo, interpretação dos parâmetros e conversão de valores entre escalas de medida. Para tornar geral, usaremos as seguintes letras maiúsculas regulares (sem itálico), Y e X , dentro de parênteses após os parâmetros para representar suas unidades de medida determinadas a partir das variáveis dependente e explicativa.

Nas subseções a seguir exemplificaremos o procedimento para reparametrização de modelos de regressão não linear frequentemente adotados em aplicações nas Ciências Agrárias. Casos envolvendo um, dois ou mais que p parâmetros serão considerados, bem como uma particular técnica de reparametrização considerada para polinômios de segundo e terceiro grau.

2.1 Reparametrização 1:1 - Modelos para acúmulo com ênfase na fração do total

O modelo Michaelis-Menten foi inicialmente proposto para para descrever a cinética de reações químicas (MICHAELIS; MENTEN, 1913). Tal modelo envolve uma função monótona crescente côncava a partir da origem. Atualmente observa-se a aplicação desse modelo em diversos contextos; um deles é a descrição do acúmulo de potássio liberado do solo (ZEVIANI et al., 2012). Sua forma funcional é

$$f(x; \theta_a, \theta_v) = \frac{\theta_a x}{\theta_v + x}, \quad x \geq 0 \text{ (X)}, \quad (2.2)$$

em que $\theta_a \geq 0$, é a assintota superior (Y) e representa o conteúdo total de nutriente liberado, e $\theta_v > 0$ é tempo de meia vida (X) ou tempo para fração meio (Figura 1).

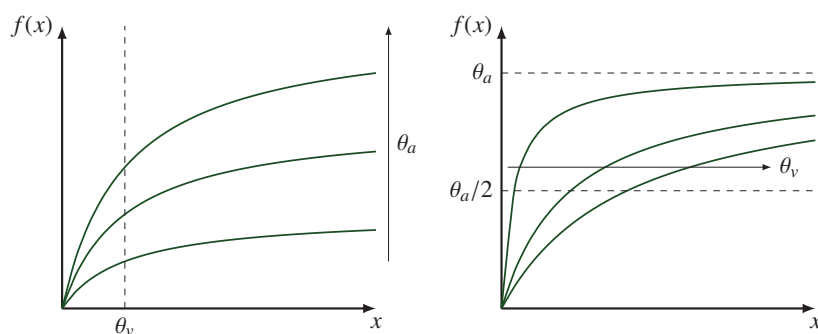


Figura 1 Modelo Michaelis-Menten com destaque para o significado e comportamento com relação aos parâmetros. A assintota superior é representado por θ_a (esq.) é o tempo de meia vida por θ_v (dir.)

Nesse modelo, os dois parâmetros têm interpretação. No entanto, algumas situações exigem conhecimento do tempo para liberação de outras frações, como 3/4 ou 95% do conteúdo total. Tal quantidade pode ser calculada a partir dos parâmetros e pela mesma razão pode ser representada no modelo. Seguindo o procedimento descrito, podemos reparametrizar o modelo para que qualquer fração

do total seja um de seus parâmetros. A fração será representada por q , $0 < q < 1$, seu valor correspondente por ϑ_q . Trata-se de uma reparametrização um para um. O valor de f no ponto $x = \vartheta_q$ é por definição igual à $q\theta_a$. Assim pode-se aplicar o primeiro passo que é expressar ϑ_q como função de θ_a e θ_v como

$$q\theta_a = \frac{\theta_a\vartheta_q}{\theta_v + \vartheta_q}, \text{ tem-se}$$

$$\vartheta_q = \frac{q}{1-q}\theta_v.$$

Então se que se $q = 1/2$, θ_v representa a fração meio. Agora aplica-se o passo 2 que é escrever algum parâmetro como função daquele de interesse. Nesse caso tem-se θ_v como única opção e então

$$\theta_v = \frac{1-q}{q}\vartheta_q.$$

Agora aplica-se o passo 3, substituindo θ_v na expressão que contém ϑ_q , obtendo

$$f(x; \theta_a, \vartheta_q) = \frac{\theta_a x}{\left(\frac{1-q}{q}\right)\vartheta_q + x}.$$

Então tem-se o modelo reparametrizado como desejava-se. Nesse modelo, q é uma constante, e não parâmetro, que identifica à fração à que $\vartheta_q > 0$ (X) corresponde.

Assim como o modelo Michaelis-Menten, o modelo exponencial assintótico também é uma função monótona crescente côncava a partir da origem. Esse modelo também tem aplicações semelhantes àquelas do modelo Michaelis-Menten, como liberação de potássio (ZEVIANI et al., 2012) além de ser usado para descrever o progresso de doenças (KRANZ, 1990; SEGARRA et al., 2001).

Sua forma funcional é

$$f(x; \theta_a, \theta_c) = \theta_a(1 - \exp\{-\theta_c x\}), \quad x \geq 0 \text{ (X)}, \quad (2.3)$$

em que $\theta_a \geq 0$ é a assíntota superior (Y) e $\theta_c > 0$ (X^{-1}) é diretamente proporcional a taxa de variação instantânea da função na origem e por isso está relacionado à velocidade de acúmulo ou progresso (Figura 2). Embora muitos trabalhos atribuam significado equivocado de taxa ao parâmetro sua dimensionalidade não é exatamente de taxa. Nesse modelo o parâmetro θ_a tem interpretação mas θ_c não possui uma interpretação prática. Da mesma forma como ocorre com o Michaelis-Menten, o tempo para a liberação de uma fração q do conteúdo total é um parâmetro de interesse. Tal forma reparametrizada do modelo foi obtida e avaliada por Zeviani et al. (2012), porém não foram dados detalhes do procedimento que agora será explorado por meio dos três passos de reparametrização. No passo 1 tem-se

$$q\theta_a = \theta_a(1 - \exp\{-\theta_c \vartheta_q\}), \text{ o que implica}$$

$$\vartheta_q = -\frac{\log(1 - q)}{\theta_c}.$$

No passo 2 escreve-se θ_c como função de ϑ_q

$$\theta_c = -\frac{\log(1 - q)}{\vartheta_q}.$$

Por fim, o passo 3 obtem-se

$$f(x; \theta_a, \vartheta_q) = \theta_a(1 - \exp\{x \log(1 - q)/\vartheta_q\}).$$

É oportuno enfatizarmos que os dois modelos apresentados possuem formas funcionais diferentes mas após reparametrização apresentam a mesma inter-

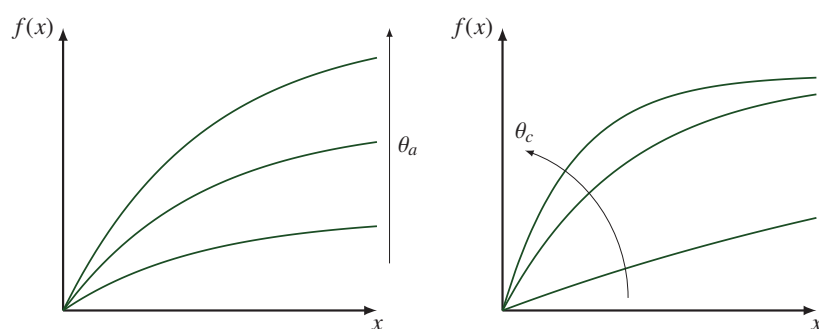


Figura 2 Modelo exponencial assintótico com destaque para o significado e comportamento com relação aos parâmetros

pretação para os parâmetros. Isso os torna diretamente comparáveis em termos de interpretação. A escolha do melhor modelo deve ser baseada na qualidade de ajuste aos dados e propriedades inferenciais conforme investigado por Zeviani et al. (2012).

2.2 Reparametrização 2:2 - Modelos com ênfase no ponto crítico

O modelo polinomial de segundo grau ou modelo quadrático é um modelo linear utilizado com frequência como uma aproximação local de uma função desconhecida. Um exemplo é sua aplicação em experimentos industriais dedicados à otimização de processos (BAS; BOYACI, 2007), em que existe interesse no ponto crítico (ou estacionário) e no valor correspondente da resposta. O modelo tem a seguinte forma funcional

$$f(x; \theta_0, \theta_1, \theta_2) = \theta_0 + \theta_1 x + \theta_2 x^2, \quad -\infty < x < \infty, \quad (2.4)$$

em que θ_0 é o intercepto (Y), θ_1 é a taxa de variação na origem (Y X⁻¹) pois $\left. \frac{\partial f(x)}{\partial x} \right|_{x=0} = \theta_1$ e θ_2 (Y X⁻²) mede o grau de curvatura e orientação da concavidade

da parábola. O espaço paramétrico de cada parâmetro é a reta real. Apesar das mencionadas interpretações que podem ter algum valor prático, o modelo quadrático é normalmente tido como empírico e sua aplicação, portanto, não é justificada pela interpretação dos parâmetros mas pelo fato de ser uma aproximação local de segunda ordem muito flexível que apresenta um ponto crítico.

Dado o interesse pelo ponto crítico ϑ_x e o valor correspondente da função ϑ_y , vamos reparametrizar o modelo para tê-los como parâmetros do modelo. Ao aplicar o passo 1 para ϑ_x tem-se que

$$\vartheta_x = \frac{-\theta_1}{2\theta_2}, \quad -\infty < \vartheta_x < \infty,$$

e conseqüentemente

$$\vartheta_y = f(\vartheta_x) = \theta_0 + \theta_1\vartheta_x + \theta_2\vartheta_x^2, \quad -\infty < \vartheta_y < \infty.$$

Considerando que θ_0 e θ_1 descrevem propriedades da função na origem ($x = 0$) e que muitas vezes a região experimental está distante da origem, usaremos θ_0 e θ_1 porque apresentam menor significado prático. No passo 2 vamos substituir θ_1 por ϑ_x e θ_0 por ϑ_y . Dessa maneira tem-se

$$\theta_1 = -2\theta_2\vartheta_x$$

$$\theta_0 = \vartheta_y - \theta_1\vartheta_x - \theta_2\vartheta_x^2.$$

Aplicando o passo 3 obtém-se

$$f(x; \theta_y, \theta_x, \theta_2) = \vartheta_y + \theta_2(x - \vartheta_x)^2, \quad -\infty < x < \infty.$$

O parâmetro ϑ_x (X) é o valor do ponto de crítico, que pode ser de máximo

ou mínimo, e ϑ_y (Y) o valor da função nesse ponto. O parâmetro θ_2 representa a diferença no valor da função com relação ao ϑ_y uma unidade acima ou abaixo do ponto ϑ_x ; ele pode ser entendido como grau de especificidade à condição ótima uma vez que quanto maior θ_2 maiores serão as diferenças da função ao se afastar do ponto crítico.

Outro modelo cujo interesse está no ponto crítico é o de Bleasdale e Nelder (1960), aplicado para representar a relação entre produção em função da densidade de plantas por unidade de área. Sua forma funcional é

$$f(x; \theta_0, \theta_1, \theta_2) = x(\theta_0 + \theta_1 x)^{-1/\theta_2}, \quad x \geq 0, \quad (2.5)$$

com parâmetros que não têm significado prático ($\theta_1 \geq 0$, θ_0 e $\theta_2 > 0$). A função apresenta ponto de máximo apenas se $\theta_2 < 1$, assim para estudos de produção-densidade, $0 < \theta_2 < 1$.

Segundo Schabenberger e Pierce (2002), a máxima produção por unidade de área, calculada por

$$\vartheta_y = \frac{\theta_2}{\theta_1} \left(\frac{1 - \theta_2}{\theta_0} \right)^{(1-\theta_2)/\theta_2} = \vartheta_y \geq 0 \text{ (Y)}, \quad (2.6)$$

que corresponde à densidade de plantas calculada por

$$\vartheta_x = \frac{\theta_0}{\theta_1} \left(\frac{\theta_2}{1 - \theta_2} \right), \quad \vartheta_x > 0 \text{ (X)}, \quad (2.7)$$

são parâmetros de fundamental importância em estudos dessa natureza. As duas expressões (2.6) e (2.7) definem os parâmetros de interesse como função dos parâmetros do modelo e portanto representam o passo 1 do procedimento de reparametrização. Pela análise dessas expressões conclui-se que é mais simples escrever

θ_0 como função de ϑ_y e θ_1 como função de ϑ_x . Se considerar o parâmetro θ_2 em alguma substituição o resultado seria uma expressão sem solução analítica. Dessa forma, para o passo 2 tem-se

$$\theta_0 = (1 - \theta_2) \left(\frac{\vartheta_y}{\vartheta_x} \right)^{-\theta_2}$$

$$\theta_1 = \frac{\theta_0}{\vartheta_x} \left(\frac{\theta_2}{1 - \theta_2} \right).$$

Finalmente, no passo 3, obtém-se após simplificações

$$f(x, \vartheta_y, \vartheta_x, \theta_2) = \vartheta_y \frac{x}{\vartheta_x} \left(1 - \theta_2 \left(1 - \frac{x}{\vartheta_x} \right) \right)^{-1/\theta_2}, \quad x \geq 0. \quad (2.8)$$

Nessa parametrização, verifica-se que o parâmetro θ_2 está relacionado a curvatura da função ao redor do ponto crítico (Figura 3) para valores fixos dos demais parâmetros. Em oposição ao modelo original, o modelo reparametrizado possui interpretação prática em ϑ_x e ϑ_y além de esclarecer que $1/\theta_2$ pode ser interpretado como coeficiente de especificidade da condição de ótimo. Tal conhecimento com relação ao modelo e seus parâmetros facilita a obtenção de valores iniciais em procedimentos de estimação, a definição de hipóteses e atribuição de prioris. Além do mais, permitem que se estude alterações fisiológicas e produtivas por meio de tais parâmetros como função, por exemplo, de materiais genéticos, de métodos de manejo, de níveis de adubação e irrigação utilizando modelos não lineares de efeitos mistos (PEEK et al., 2002).

2.3 Excesso de parâmetros de interesse

O modelo de Wood (1967) é muito utilizado para descrever a curva de lactação de vacas após o parto (OLORI et al., 1999). Sua forma funcional é baseada

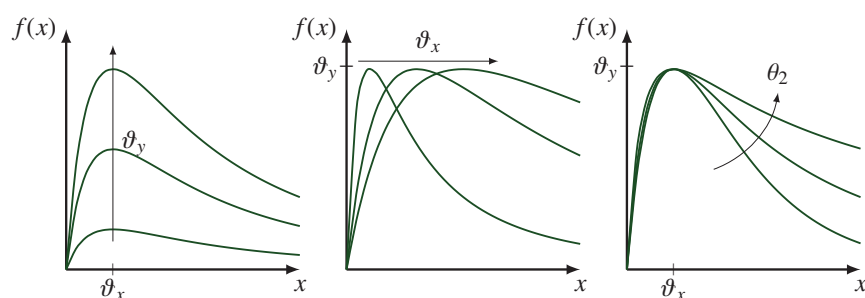


Figura 3 Modelo de Bleasdale-Nelder reparametrizado para o ponto crítico com destaque para o significado e comportamento com relação aos parâmetros

no núcleo da distribuição de probabilidades Gama e por isso também é chamado de modelo gama incompleto (CONGLETON; EVERETT, 1980). Geralmente ele é expresso como

$$f(x) = \theta_0 x^{\theta_1} \exp\{-\theta_2 x\}, \quad x \geq 0, \quad (2.9)$$

em que $\theta_0 > 0$ ($Y X^{\theta_1}$) é um fator de escala, $\theta_1 > 0$ está associado à taxa de variação antes do pico de produção e $\theta_2 > 0$ (X^{-1}) a taxa de variação após o pico de produção de leite. Segundo Guimarães et al. (2006) são de interesse prático três parâmetros obtidos como função destes. São eles

$$\vartheta_y = \theta_0 (\theta_1 / \theta_2)^{\theta_1} \exp\{-\theta_1\}, \quad \vartheta_y > 0 \text{ (Y)} \quad (2.10)$$

que representa máxima quantidade produzida de leite,

$$\vartheta_x = \theta_1 / \theta_2, \quad \vartheta_x > 0 \text{ (X)} \quad (2.11)$$

que corresponde ao número de dias necessários para atingir o máximo de produção, e

$$\vartheta_p = \theta_2^{-(\theta_1+1)}, \quad \vartheta_p > 0 \quad (\mathbf{X}^{\theta_1+1}) \quad (2.12)$$

que está relacionado à persistência na produção pois é proporcional ao intervalo de tempo em que a vaca sustenta a produção próxima ao valor máximo.

Além destes, outro parâmetro útil, associado ao potencial produtivo da vaca, é a área abaixo da curva de lactação (FRANCE; THORNLEY, 2006). Considerando um tempo de lactação infinito, a área abaixo da curva é

$$\vartheta_a = \frac{\theta_0 \Gamma(\theta_1 + 1)}{\theta_2^{\theta_1+1}}, \quad \vartheta_a > 0 \quad (\mathbf{Y} \mathbf{X}). \quad (2.13)$$

Este é o caso em que o número de parâmetros de interesse é maior que o número de parâmetros do modelo. Obviamente, a reparametrização, nesse caso, poderá representar no máximo três desses parâmetros de interesse. Algumas das possíveis funções desses parâmetros são

$$\begin{aligned} \theta_0 &= \vartheta_y \left(\frac{1}{\vartheta_x} \right)^{\theta_1} \exp\{\theta_1\} = \frac{\vartheta_a \theta_2^{\theta_1+1}}{\Gamma(\theta_1 + 1)} \\ \theta_1 &= \vartheta_x \theta_2 = \frac{\log(\vartheta_p)}{\log(\theta_2)} - 1 \\ \theta_2 &= \frac{\theta_1}{\vartheta_x} = \vartheta_p^{-1/(\theta_1+1)}. \end{aligned}$$

Não é possível incluir simultaneamente três desses parâmetros de interesse na reparametrização de forma puramente analítica. Se se considerar pares de parâmetros, obtem-se reparametrizações que contêm qualquer um dos quatro

parâmetros de interesse. Um resultado possível é o que inclui ϑ_y e ϑ_x ,

$$f(x; \vartheta_y, \vartheta_x, \theta_1) = \vartheta_y \left(\frac{x}{\vartheta_x} \right)^{\theta_1} \exp\{\theta_1(1 - x/\vartheta_x)\}. \quad (2.14)$$

Neste caso, o parâmetro θ_1 têm relação com a curvatura ao redor do ponto de máximo da função (Figura 4). Outra possibilidade é o que inclui ϑ_p e ϑ_a ,

$$f(x; \vartheta_p, \vartheta_a, \theta_1) = \frac{\vartheta_a}{\vartheta_p \Gamma(\theta_1 + 1)} x^{\theta_1} \exp\{-x \vartheta_p^{-1/(\theta_1+1)}\}. \quad (2.15)$$

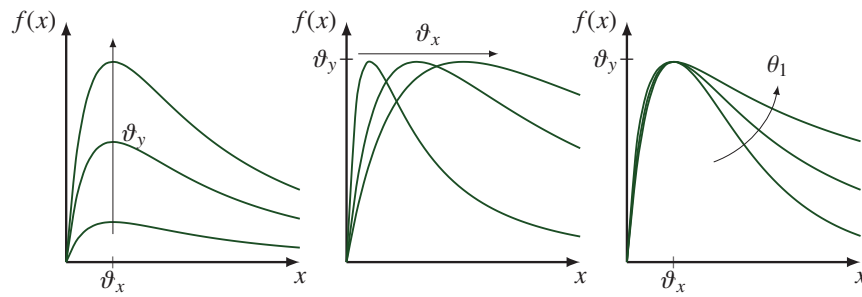


Figura 4 Modelo de Wood reparametrizado para o ponto crítico com destaque para o significado e comportamento com relação aos parâmetros

2.4 Uso de soluções numéricas

Na seção anterior, concluímos não ser possível incorporar em (2.9) três dos parâmetros de interesse mediante soluções analíticas. No entanto, nesse modelo em particular, isso é possível por meio de soluções numéricas. A dificuldade analítica encontrada é a de substituir θ_1 por ϑ_p ou ϑ_a . Considerando ϑ_p como

exemplo, temos a relação com θ_1 ,

$$\begin{aligned}\theta_1 &= \vartheta_x \vartheta_p^{-1/(\theta_1+1)} \Rightarrow \\ \theta_1 - \vartheta_x \vartheta_p^{-1/(\theta_1+1)} &= 0,\end{aligned}$$

e essa última equação pode ser resolvida, para θ_1 com ϑ_x e ϑ_p conhecidos, usando qualquer algoritmo, como o Newton-Raphson. Com isso podemos reparametrizar o modelo de Wood de forma a representar ϑ_y , ϑ_x e ϑ_p . No entanto, o modelo não terá uma forma funcional fechada mas uma forma funcional que depende de um passo numérico,

$$f(x; \vartheta_y, \vartheta_x, \vartheta_p) = \vartheta_y \left(\frac{x}{\vartheta_x} \right)^{\hat{\theta}_1} \exp\{\hat{\theta}_1(1 - x/\vartheta_x)\},$$

em que $\hat{\theta}_1$ é a solução de $\hat{\theta}_1 - \vartheta_x \vartheta_p^{-1/(\hat{\theta}_1+1)} = 0$.

Para avaliar a verossimilhança do modelo para o vetor $\boldsymbol{\theta} = (\vartheta_y, \vartheta_x, \vartheta_p)^\top$ é necessário um passo numérico para determinar $\hat{\theta}_1$ a partir de ϑ_x e ϑ_p . Em termos de implementação em ambiente R, pode-se ajustar tal modelo por meio da função `bbmle::mle2()` e no passo para determinar $\hat{\theta}_1$ considerar a função `rootSolve::uniroot()`.

2.5 Ponto crítico em polinômios de grau 2 e 3

A reparametrização do modelo quadrático (polinômio de grau 2) também pode ser obtida por uma aproximação em série de Taylor de segunda ordem em torno do ponto crítico ϑ_x . Ou seja, nesse caso não se trata-se de uma aproximação uma vez que é exata, mas de uma translação da origem para o ponto de inflexão.

Por definição, a primeira derivada da função no ponto crítico ϑ_x é zero e o valor da função é ϑ_y . Dessa forma

$$\begin{aligned} f(x) &= f(\vartheta_x) + f'(\vartheta_x)(x - \vartheta_x) + \frac{f''(\vartheta_x)}{2}(x - \vartheta_x)^2 \\ &= \vartheta_y + \theta_2(x - \vartheta_x)^2. \end{aligned}$$

Assim como o modelo quadrático, o modelo cúbico é um modelo linear empregado por vezes como uma aproximação local desprovido de interpretação prática para seus parâmetros. Por se tratar de um polinômio, esse modelo é muito flexível e é por isso considerado como aproximação de alguns modelos sigmóides e do modelo beta usado para descrever respostas biológicas como função da temperatura (YIN et al., 1995), no qual é de interesse determinar a temperatura de ótimo (ϑ_x) e valor correspondente da resposta (ϑ_y).

Usando a expansão em série de Taylor do modelo cúbico, pode-se reparametrizá-lo em termos do ponto crítico. Dessa forma tem-se

$$\begin{aligned} f(x) &= \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \\ f(x) &= f(\vartheta_x) + f'(\vartheta_x)(x - \vartheta_x) + \frac{f''(\vartheta_x)}{2}(x - \vartheta_x)^2 + \frac{f'''(\vartheta_x)}{6}(x - \vartheta_x)^3 \\ &= \vartheta_y + \theta_2^*(x - \vartheta_x)^2 + \theta_3(x - \vartheta_x)^3, \end{aligned}$$

em que ϑ_y (Y) é o valor da função no ponto crítico ϑ_x (X), os parâmetros θ_2^* (Y X⁻²) e θ_3 (Y X⁻³) controlam a concavidade e assimetria da função (Figura 5).

A expansão em série de Taylor pode ser considerada para reparametrizar polinômios de grau maior que três com relação ao ponto crítico, porém tais polinômios são de baixa aplicabilidade, além de apresentarem muitos pontos críticos.

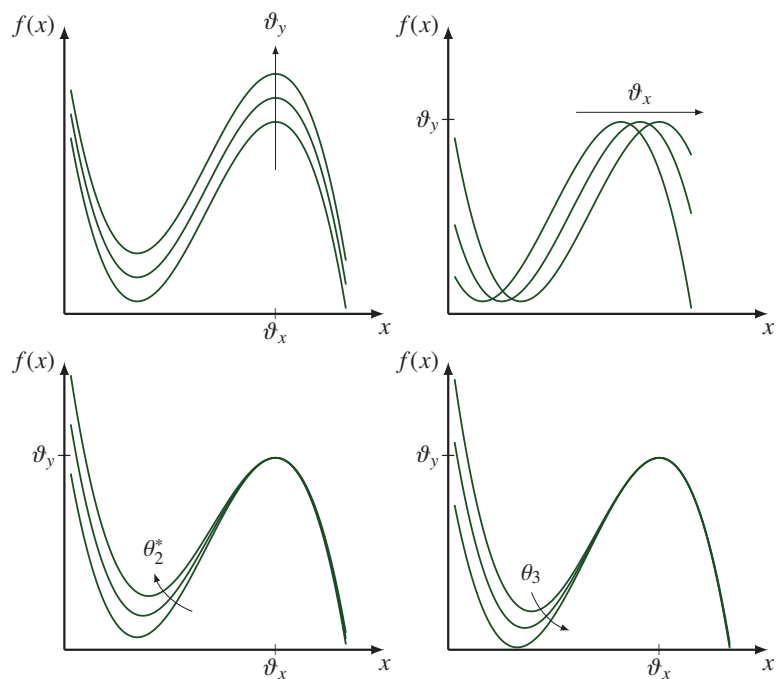


Figura 5 Modelo de cúbico reparametrizado para o ponto crítico com destaque para o significado e comportamento com relação aos parâmetros

2.6 Distribuições de probabilidade

Em geral, nos modelos de regressão, o interesse é modelar a média de uma variável aleatória Y como função de covariáveis. Na distribuição Normal, o parâmetro μ , que representa a média da distribuição de probabilidades, é parâmetro da função densidade de probabilidade e modelos de regressão são desenvolvidos para avaliar associação de variáveis dependentes à média de Y . As funções densidade de probabilidade são tipicamente não lineares em seus parâmetros. Por isso, ao assumir uma distribuição não Normal para a variável dependente, como Gama, Beta, Weibull, surge a preocupação de como ligar o preditor à média que não é parâmetro da função densidade. Haja visto que a média pode ser obtida por função dos

parâmetros da função densidade, por meio de reparametrização é possível então representá-la.

Considere a função densidade da distribuição Gama na forma

$$f(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp\{-\beta y\}, \quad y > 0. \quad (2.16)$$

Objetiva-se representar com parâmetros, diretamente na função densidade, a média e a variância. Nessa parametrização $E(Y) = \mu = \alpha/\beta$ e $\text{Var}(Y) = \sigma^2 = \alpha/\beta^2$. Assim, temos que $\beta = \mu/\sigma^2$ e $\alpha = \mu^2/\sigma^2$. Substituindo essas expressões, tem-se que a função densidade da Gama parametrizada para média e variância é escrita como

$$f(y; \mu, \sigma^2) = \frac{(\mu/\sigma^2)^{\mu^2/\sigma^2}}{\Gamma(\mu^2/\sigma^2)} y^{\mu^2/\sigma^2-1} \exp\{-y\mu/\sigma^2\}, \quad y > 0. \quad (2.17)$$

Bonat et al. (2012) avaliaram diferentes parametrizações da distribuição Gama por meio da visualização da função de log-verossimilhança e concluíram que a parametrização da média, além de ser diretamente interpretável, tem função de log-verossimilhança mais simétrica com semieixos praticamente ortogonais entre si com orientação paralela aos eixos dos parâmetros.

Da mesma forma, para a distribuição Beta

$$f(y; \alpha_1, \alpha_2) = \frac{1}{B(\alpha_1, \alpha_2)} y^{\alpha_1-1} (1-y)^{\alpha_2-1}, \quad 0 < y < 1, \quad (2.18)$$

em que B é a função beta, temos que $\mu = \alpha_1/(\alpha_1 + \alpha_2)$ e $\sigma^2 = \alpha_1\alpha_2/((\alpha_1 + \alpha_2 + 1)(\alpha_1 + \alpha_2)^2)$. Podemos escrever $\alpha_1 = \mu\alpha_2/(1 - \mu)$. Escrever α_2 como função de σ resulta em uma expressão complicada e não útil na prática pois nos modelos de regressão os parâmetros de dispersão não são alvo de modelagem. Ao parametrizar

(2.18) apenas com relação à média, obtemos

$$f(y; \mu, \alpha_2) = \frac{1}{B(\mu\alpha_2/(1-\mu), \alpha_2)} y^{\mu\alpha_2/(1-\mu)-1} (1-y)^{\alpha_2-1}, \quad 0 < y < 1, \quad (2.19)$$

em que o parâmetro α_2 pode ser entendido como parâmetro de precisão. Ferrari e Cribari-Neto (2004) apresentam uma parametrização diferente desta no parâmetro de precisão.

Ambos os modelos têm a média representada na função densidade e para fins de modelagem estatística, um preditor, normalmente linear, pode ser diretamente associado a este parâmetro.

2.7 Outros modelos

Além dos modelos considerados para exemplificar o procedimento de reparametrização, outros modelos frequentemente aplicados em Ciências Agrárias foram reparametrizados e estão apresentados nas tabelas de 1 à 3. A descrição de cada modelo, em termos de propriedades da função não linear, interpretação dos parâmetros é fornecida a seguir em uma lista numerada de acordo com as linhas das tabelas de 1 a 3.

Para uma descrição genérica dos modelos e adequada representação da dimensionalidade dos parâmetros e demais grandezas, a seguinte notação será usada: A letra X representa a unidade de medida da variável independente x e Y representa a unidade de medida da variável dependente. Quando unidade de medida não for apresentada considere a grandeza adimensional. Quando o suporte ou espaço paramétrico não for definido considere a reta real.

1. O modelo Michaelis-Menten (MICHAELIS; MENTEN, 1913), proveniente da cinética de reações químicas envolve uma função é não negativa, monó-

tona crescente e côncava considerando que $x > 0$. $\theta_a \geq 0$ (Y) é assíntota ($\lim_{x \rightarrow \infty} f(x) = \theta_a$) e $\theta_v > 0$ (X) é o tempo de meia vida, que corresponde à abscissa para qual o valor da função é $\theta_a/2$. No modelo reparametrizado, $0 < q < 1$ é uma constante que representa uma fração de θ_a para o qual $\vartheta_q > 0$ (X) é o valor correspondente na abscissa. Então ϑ_q é o tempo para uma fração q de vida. Zeviani et al. (2012) consideraram tal modelo para estudar a liberação de potássio no solo a partir de fontes orgânicas incubadas em latossolos.

2. É uma extensão do modelo Michaelis-Menten (item 1) pela reparametrização e introdução do parâmetro θ_c . A função é positiva e monótona crescente considerando $x \geq 0$, $\theta_a \geq 0$ (Y) é a assíntota, θ_v (X) é o tempo de meia vida e $\theta_c > 0$ controla a forma da função. Se $0 < \theta_c < 1$ a função é côncava, se $\theta_c > 1$ têm formato sigmóide, ou seja, apresenta ponto de inflexão, e se $\theta_c = 1$ reduz-se ao modelo Michaelis-Menten, numa parametrização diferente da original. No modelo reparametrizado, $0 < q < 1$ é uma constante que representa uma fração de θ_a para o qual $\vartheta_q > 0$ (X) é o valor correspondente na abscissa. Então ϑ_q é o tempo para uma fração q de vida. Groot et al. (1996) consideraram esse modelo para representar a produção de gases em ruminantes e Becker et al. (2007) para estudar o crescimento de bactérias como função do tempo.
3. Uma modificação do modelo do item 2 no qual a função passa ser monótona decrescente considerando que $\theta_a > 0$ (Y) é agora intercepto ($f(x = 0) = \theta_a$) e que a assíntota inferior é o zero ($\lim_{x \rightarrow \infty} f(x) = 0$). O parâmetro $\vartheta_q > 0$ (X) representa o valor na abscissa correspondente a uma fração q de vida.
4. Uma extensão do modelo Michaelis-Menten pela adição do parâmetro θ_c

conhecido por modelo de Hill (HILL, 1913). A função é positiva e monótona crescente considerando $x \geq 0$, $\theta_a \geq 0$ (Y) é a assíntota, θ_v^{1/θ_c} (X) é o tempo de meia vida e $\theta_c > 0$ controla a forma da função. Se $0 < \theta_c < 1$ a função é côncava, se $\theta_c > 1$ têm formato sigmoide, e se $\theta_c = 1$ reduz-se ao modelo Michalis-Menten. Goutelle et al. (2008) descrevem várias propriedades do modelo de Hill.

5. Conhecido por modelo exponencial assintótico (PINHEIRO; BATES, 2000), negativo exponencial (TIORVE, 2003) ou monomolecular (BOX; LUCAS, 1959) é uma função positiva, monótona crescente e côncava considerando $x \geq 0$, $\theta_a > 0$ é a assíntota e θ_c (X^{-1}) está associado à taxa. No modelo reparametrizado, $0 < q < 1$ é uma constante que representa uma fração de θ_a para o qual $\vartheta_q > 0$ (X) é o valor correspondente na abscissa, ou seja, é o tempo para uma fração q de vida. Zeviani et al. (2012) consideraram a versão reparametrizada desse modelo para estudar a liberação de potássio no solo a partir de fontes orgânicas incubadas em latossolos.
6. Uma extensão do modelo exponencial assintótico (item 5) pela adição do parâmetro θ_0 (PINHEIRO; BATES, 2000) também conhecido como modelo Mitscherlich (SCHABENBERGER; PIERCE, 2002). A função é positiva, monótona crescente e côncava considerando $x \geq 0$, $\theta_a > 0$ é a assíntota, θ_1 (X^{-1}) está associado à taxa e θ_0 (X) o tempo de aquecimento pois a função é nula para $x < \theta_0$. No modelo reparametrizado, $0 < q < 1$ é uma constante que representa uma fração de θ_a para o qual $\vartheta_q > 0$ (X) é o valor correspondente na abscissa, ou seja, é o tempo para uma fração q de vida. Peek et al. (2002) consideram tal modelo para avaliar o acúmulo de CO_2 fotossintético como função da intensidade luminosa em plantas em diferentes ambientes.

7. Conhecido como modelo de Herschel-Bulkley (CHENG et al., 1992) é uma extensão do modelo linear simples (equação da reta) pela adição do parâmetro θ_2 . A função é monótona decrescente considerando $x \geq 0$, θ_0 (Y) é o intercepto, $\theta_1 \geq 0$ ($X^{-\theta_2}$) corresponde à redução no valor do intercepto para $x = 1$, ou seja, $\theta_1 = f(0) - f(1)$ e $\theta_2 > 0$ é o parâmetro que controla a concavidade. Se $0 < \theta_2 < 1$ a função côncava, se $\theta_2 > 1$ é convexa e se $\theta_2 = 1$ reduz-se à equação da reta. No modelo reparametrizado, $q > 0$ (Y) é uma constante que representa uma redução em θ_0 para o qual $\vartheta_q > 0$ (X) é o valor correspondente na abscissa, ou seja, $\theta_0 - q = f(0) - f(\vartheta_q)$. Tal modelo pode ser considerado para modelar respostas assintóticas crescentes e decrescentes como a produção em função de um estímulo, por exemplo a adubação ou irrigação, ou em função de um estresse, por exemplo, o ataque de uma praga/doença ou déficit hídrico.
8. Uma parametrização alternativa do modelo assintótico $f(x) = \theta_a - \exp\{-\theta_b\}\theta_c^x$ apresentado por Ratkowsky (1983). A função é monótona e muito flexível pois pode ser côncava crescente ($0 < \theta_c < 1$ e $\theta_1 < 0$), côncava decrescente ($\theta_c > 1$ e $\theta_1 < 0$), convexa crescente ($\theta_c > 1$ e $\theta_1 > 0$) e convexa decrescente ($0 < \theta_c < 1$ e $\theta_1 > 0$) considerando $x \geq 0$, θ_0 (Y) é o intercepto, $\theta_c > 0$ e θ_1 não interpretáveis. No modelo reparametrizado, $q > 0$ (Y) é uma constante que representa uma redução em θ_0 para o qual $\vartheta_q > 0$ (X) é o valor correspondente na abscissa, ou seja, $\theta_0 - q = f(0) - f(\vartheta_q)$. Tal modelo pode ser considerado para as mesmas situações descritas para o modelo de Herschel-Bulkley (item 7).
9. É um modelo segmentado conhecido como linear-platô pois é a junção de dois segmentos de reta no qual um deles é horizontal (SCHABENBERGER; PIERCE, 2002). A função é crescente (decrescente) para $\theta_1 > 0$ ($\theta_1 < 0$) até

o valor $0 < x \leq \theta_b$, e para $x > \theta_b$ a função é constante. A derivada da função em x é descontínua no ponto θ_b pois tem valor θ_1 para $x \leq \theta_b$ e 0 para $x > \theta_b$. O intercepto é representado por θ_0 (Y) e θ_1 ($Y X^{-1}$) corresponde à taxa de variação antes do ponto de quebra $\theta_b > 0$ (X). No modelo reparametrizado, ϑ_b (Y) representa o valor da função no ponto de quebra θ_b que é o máximo (mínimo) valor da função se $\theta_1 > 0$ ($\theta_1 < 0$). No contexto de estímulo em função de doses, como produção em função da adubação, θ_b é interpretado como dose econômica e ϑ_b como produção máxima. Além disso, esse modelo pode ser considerado para as mesmas situações descritas para o modelo de Herschel-Bulkley (item 7). Carneiro et al. (2011) (2011) o consideraram para estudar tamanho ótimo de parcela para avaliação da resistência do feijoeiro ao mofo branco e Peixoto et al. (2011) para conservação *in vitro* de maracujazeiro.

10. É um modelo segmentado conhecido como bi-linear segmentado pois é a junção de dois segmentos de reta (VIETH, 1989). A função assume diversas formas dependendo dos valores dos parâmetros. O intercepto é representado por θ_0 (Y), θ_1 e θ_2 ($Y X^{-1}$) são os coeficientes ou taxas do primeiro e segundo segmentos de reta, respectivamente, que se encontram no ponto θ_b (X), denominado de ponto de quebra. A derivada da função é descontínua em θ_b . No modelo reparametrizado, o ponto de quebra foi priorizado e representado de tal modo que ϑ_b (Y) é o valor da função no ponto de quebra θ_b . Se $\theta_2 = 0$ o modelo reduz-se ao linear-platô (item 9), se $\theta_1 = 0$ reduz-se ao modelo platô-linear e se $\theta_1 = \theta_2$ reduz-se ao modelo linear simples. Os modelos lineares segmentados podem ter mais de um ponto de quebra conforme Muggeo (2003). Shuai et al. (2003) considerou essa classe de modelos para modelar o rendimento de grãos de soja e o teor de nutrientes

como função do pH do solo controlado pela aplicação de calcário.

11. Trata-se do polinômio de segundo grau, um modelo linear nos parâmetros, chamado também de modelo quadrático. A função assume diversas formas dependendo dos valores dos parâmetros e apresenta um ponto crítico (de máximo ou mínimo) e é simétrica com relação a este. O intercepto é θ_0 (Y), θ_1 ($Y X^{-1}$) é coeficiente angular que representa a taxa na origem, θ_2 ($Y X^{-2}$) é a curvatura. No modelo reparametrizado, o ponto crítico foi priorizado e representado de tal modo que ϑ_y (Y) é o valor no ponto crítico, que será de máximo se $\theta_2 < 0$ e mínimo se $\theta_2 > 0$, correspondente ao valor ϑ_x (Y) na abscissa. O parâmetro θ_2 pode ser interpretado como coeficiente de especificidade uma vez que mede a intensidade de mudança no valor da função ao se afastar de ϑ_x . Nessa parametrização, o modelo é conhecido forma canônica do modelo quadrático (KHURI, 2006). Bullock e Bullock (1994) considerou esse modelo para estabelecer recomendações de adubação nitrogenada para a cultura do milho e Serafim et al. (2012) avaliaram o efeito de doses de potássio na produtividade de soja em associação à níveis de déficit hídrico.
12. É um modelo segmentado conhecido como modelo quadrático-platô pois é junção de reta horizontal com o ponto crítico do modelo quadrático. A função pode ser crescente ou decrescente na porção representada pelo modelo quadrático e constante para $x > \theta_b$. O intercepto é θ_0 (Y), θ_1 ($Y X^{-1}$) é coeficiente angular que representa a taxa na origem, θ_2 ($Y X^{-2}$) é a curvatura. A junção das funções é no ponto crítico do modelo quadrático calculado por $-\theta_1/(2\theta_2)$ e apresenta derivada contínua. No modelo reparametrizado, o ponto crítico foi priorizado e representado de tal modo que ϑ_y (Y) é o valor máximo ($\theta_2 < 0$) ou mínimo ($\theta_2 > 0$) da função no ponto de junção ϑ_x (X) que em estudos de adubação é interpretado como dose econômica (BOCK;

SIKORA, 1990). Carneiro et al. 2011 (2011) o consideraram para estudar tamanho ótimo de parcela para avaliação da resistência do feijoeiro ao mofo branco e Peixoto et al. (2011) para conservação *in vitro* de maracujazeiro.

13. Modelo de Bleasdale-Nelder (BLEASDALE; NELDER, 1960) considerado para estimar o número de plantas para maximizar a produtividade por unidade de área. A função é positiva, monótona crescente e côncava para $\theta_2 > 1$ e positiva com ponto de máximo para $0 < \theta_2 < 1$, em todos os casos $x \geq 0$ e $\theta_0 + \theta_1 x > 0$. Os parâmetros não tem interpretação prática. Quando $\theta_2 = 1$ reduz-se ao modelo de Shinozaki-Kira (SCHABENBERGER; PIERCE, 2002). No modelo reparametrizado, o ponto crítico foi priorizado e representado de tal modo que $\vartheta_y > 0$ (Y) é o valor máximo correspondente ao valor na abscissa $\theta_x > 0$ (X) e $0 < \theta_2 < 1$ está associado à curvatura da função ao redor do máximo. Watkinson (1980) apresentou parametrizações alternativas desse modelo e Damgaard e Borksted (2004) o consideraram para avaliar plantas selvagens e transgênicas de *Arabidopsis thaliana* com relação à produção de sementes por unidade de área.
14. Conhecido como modelo de (WOOD, 1967) ou modelo gama incompleto (CONGLETON; EVERETT, 1980) é frequentemente aplicado em estudos de curvas de lactação. A função é positiva e apresenta um ponto de máximo, $x > 0$. Os parâmetros $\theta_0 > 0$ ($Y X^{\theta_1}$), θ_1 e $\theta_2 > 0$ (X^{-1}) não têm interpretação, prática. No modelo reparametrizado, ênfase foi dada à parâmetros de interpretação no contexto de lactação, como medidas descritivas do ponto de máximo e a persistência. Assim sendo, $\vartheta_y > 0$ (Y) é o valor máximo correspondente ao valor na abscissa $\theta_x > 0$ (X) e ϑ_p é o parâmetro de persistência na lactação, relacionado à curvatura da função ao redor do máximo. Rekaya et al. (2001) consideraram uma parametrização desse modelo que inclui o

parâmetro de persistência e avaliaram, por meio de um modelo não linear hierárquico bayesiano, o efeito de ordem de lactação, da variação entre e dentro de vacas além de incorporar a informação de parentesco.

15. Modelo logístico aplicado para descrever curvas de crescimento individual, populacional (PINHEIRO; BATES, 2000; SCHABENBERGER; PIERCE, 2002) e progresso de doenças (SEGARRA et al., 2001). A função é sigmóide e simétrica com relação ao ponto de inflexão. O parâmetro θ_a (Y) é a assíntota superior do modelo, θ_0 e θ_1 (X^{-1}) não têm interpretação direta. No modelo reparametrizado, representou-se a taxa máxima de crescimento θ_t (X^{-1}) relativo à θ_a , ou seja, a taxa máxima absoluta ($Y X^{-1}$) dividida pela assíntota ou tamanho máximo θ_a e o tempo ϑ_q (X) para alcançar uma fração $0 < q < 1$ do tamanho final.
16. Conhecido como modelo de Gompertz, assim como o modelo logístico (item 15), é aplicado para descrever curvas de crescimento e progresso de doença. A função é sigmoide e não simétrica com relação ao ponto de inflexão. O parâmetro θ_a (Y) é a assíntota superior do modelo, θ_0 e θ_1 (X^{-1}) não têm interpretação direta. No modelo reparametrizado, representou-se o tempo ϑ_q (X) para alcançar uma fração $0 < q < 1$ do tamanho final.
17. É uma parametrização alternativa apresentada por Omuto et al. (2006) do modelo van Genuchten (GENUCHTEN, 1980) usado para descrever a curva de retenção de água do solo. A função é sigmoide e não simétrica com relação ao ponto de inflexão. As assíntotas superior e inferior são apresentadas por $\theta_s > 0$ (Y) e $\theta_r > 0$ (Y), θ_a (X) está associado com a posição do ponto de inflexão, $\theta_n > 0$ (X^{-1}) e $\theta_m > 0$ são parâmetros sem interpretação, x é o logaritmo da tensão matricial. No modelo reparametrizado, parâmetros que

descrevem o ponto de inflexão foram representados de modo que $\vartheta_i(X)$ é o ponto de inflexão e $\vartheta_s(Y X^{-1})$ é a taxa no ponto de inflexão. (DEXTER, 2004) aponta que ϑ_s têm alta associação com medidas de qualidade física do solo.

Tabela 1 Reparameatrizações desenvolvidas com ênfase na interpretação dos parâmetros de modelos de regressão não linear aplicados em Ciências Agrárias

id	Modelo original	$\vartheta = g(\theta)$	$\theta_i = g^{-1}(\vartheta, \theta_{-i})$	Modelo reparametrizado
1	$\frac{\theta_a x}{\theta_v + x}$	$\vartheta_q = \theta_v \left(\frac{q}{1-q} \right)$	$\theta_v = \vartheta_q \left(\frac{1-q}{q} \right)$	$\frac{\theta_a x}{\vartheta_q \left(\frac{1-q}{q} \right) + x}$
2	$\frac{\theta_a}{1 + \left(\frac{\theta_c}{x} \right)^{\theta_c}}$	$\vartheta_q = \theta_v \left(\frac{1-q}{q} \right)^{-1/\theta_c}$	$\theta_v = \vartheta_q \left(\frac{1-q}{q} \right)^{1/\theta_c}$	$\frac{\theta_a}{1 + \frac{1-q}{q} \left(\frac{\vartheta_q}{x} \right)^{\theta_c}}$
3	$\frac{\theta_a}{1 + \left(\frac{x}{\theta_c} \right)^{\theta_c}}$	$\vartheta_q = \theta_v \left(\frac{1-q}{q} \right)^{1/\theta_c}$	$\theta_v = \vartheta_q \left(\frac{1-q}{q} \right)^{-1/\theta_c}$	$\frac{\theta_a}{1 + \frac{1-q}{q} \left(\frac{x}{\vartheta_q} \right)^{\theta_c}}$
4	$\frac{\theta_a x^{\theta_c}}{\theta_v + x^{\theta_c}}$	$\vartheta_q = \left(\frac{\theta_v q}{1-q} \right)^{1/\theta_c}$	$\theta_v = \vartheta_q^{\theta_c} \frac{1-q}{q}$	$\frac{\theta_a x^{\theta_c}}{\vartheta_q \left(\frac{1-q}{q} \right) + x^{\theta_c}}$
5	$\theta_a(1 - \exp\{-\theta_c x\})$	$\vartheta_q = -\frac{\log(1-q)}{\theta_c}$	$\theta_c = -\frac{\log(1-q)}{\vartheta_q}$	$\theta_a(1 - \exp\{x \log(1-q)/\vartheta_q\})$
6	$\begin{cases} \theta_a(1 - \exp\{-\theta_1(x - \theta_0)\}) & , x \geq \theta_0 \\ 0 & , x < \theta_0 \end{cases}$	$\vartheta_q = \frac{\log(1-q)}{\theta_1} + \theta_0$	$\theta_1 = \frac{\log(1-q)}{\vartheta_q - \theta_0}$	$\theta_a \left(1 - \exp \left\{ \log(1-q) \left(\frac{x - \theta_0}{\vartheta_q - \theta_0} \right) \right\} \right)$
7	$\theta_0 - \theta_1 x^{\theta_2}$	$\vartheta_q = \frac{q^{1/\theta_2}}{\theta_1}$	$\theta_2 = \frac{\log(q) - \log(\theta_1)}{\log(\vartheta_q)}$	$\theta_0 - \theta_1 x^{\frac{\log(q) - \log(\theta_1)}{\log(\vartheta_q)}}$
8	$\theta_0 + \theta_1(1 - \theta_c^x)$	$\vartheta_q = \frac{\log(1 + q/\theta_1)}{\log(\theta_c)}$	$\theta_1 = -\frac{q}{1 - \theta_c^{\vartheta_q}}$	$\theta_0 - q \left(\frac{1 - \theta_c^x}{1 - \theta_c^{\vartheta_q}} \right)$

Tabela 2 (cont.) Reparametrizações desenvolvidas com ênfase na interpretação dos parâmetros de modelos de regressão não linear aplicados em Ciências Agrárias.

id	Modelo original	$\vartheta = g(\theta)$	$\theta_i = g^{-1}(\vartheta, \theta_{-i})$	Modelo reparametrizado
9	$\begin{cases} \theta_0 + \theta_1 x & , x \leq \theta_b \\ \theta_0 + \theta_1 \theta_b & , x > \theta_b \end{cases}$	$\vartheta_b = \theta_0 + \theta_1 \theta_b$	$\theta_0 = \vartheta_b - \theta_1 \theta_b$	$\begin{cases} \vartheta_b + \theta_1(x - \theta_b) & , x \leq \theta_b \\ \vartheta_b & , x > \theta_b \end{cases}$
10	$\begin{cases} \theta_0 + \theta_1 x & , x \leq \theta_b \\ \theta_0 + \theta_1 \theta_b + \theta_2(x - \theta_b) & , x > \theta_b \end{cases}$	$\vartheta_b = \theta_0 + \theta_1 \theta_b$	$\theta_0 = \vartheta_b - \theta_1 \theta_b$	$\begin{cases} \vartheta_b + \theta_1(x - \theta_b) & , x \leq \theta_b \\ \vartheta_b + \theta_2(x - \theta_b) & , x > \theta_b \end{cases}$
11	$\theta_0 + \theta_1 x + \theta_2 x^2$	$\vartheta_x = -\frac{\theta_1}{2\theta_2}$ $\vartheta_y = \theta_0 + \theta_1 \vartheta_x + \theta_2 \vartheta_x^2$	$\theta_1 = 2\theta_2 \vartheta_x$ $\theta_0 = \vartheta_y - \theta_1 \vartheta_x - \theta_2 \vartheta_x^2$	$\vartheta_y + \theta_2(x - \vartheta_x)^2$
12	$\begin{cases} \theta_0 + \theta_1 x + \theta_2 x^2, & x \leq -\theta_1/(2\theta_2) \\ \theta_0 + \theta_1 \left(\frac{-\theta_1}{2\theta_2}\right) + \theta_2 \left(\frac{-\theta_1}{2\theta_2}\right)^2, & x > -\theta_1/(2\theta_2) \end{cases}$	$\vartheta_x = -\frac{\theta_1}{2\theta_2}$ $\vartheta_y = \theta_0 + \theta_1 \vartheta_x + \theta_2 \vartheta_x^2$	$\theta_1 = 2\theta_2 \vartheta_x$ $\theta_0 = \vartheta_y - \theta_1 \vartheta_x - \theta_2 \vartheta_x^2$	$\begin{cases} \vartheta_y + \theta_2(x - \vartheta_x)^2 & , x \leq \vartheta_x \\ \vartheta_y & , x > \vartheta_x \end{cases}$
13	$x(\theta_0 + \theta_1 x)^{-1/\theta_2}$	$\vartheta_x = \frac{\theta_0}{\theta_1} \left(\frac{\theta_2}{1 - \theta_2}\right)^{1/\theta_2}$ $\vartheta_y = \vartheta_x \left(\frac{1 - \theta_2}{\theta_0}\right)^{1/\theta_2}$ $\vartheta_x = \theta_1 / \theta_2$	$\theta_1 = \frac{\theta_0}{\vartheta_x} \left(\frac{\theta_2}{-1\theta_2}\right)^{-\theta_2}$ $\theta_0 = (1 - \theta_2) \left(\frac{\vartheta_y}{\vartheta_x}\right)^{-\theta_2}$ $\theta_1 = \theta_2 \vartheta_x$	$\vartheta_y \frac{x}{\vartheta_x} \left(1 - \theta_2 \left(1 - \frac{x}{\vartheta_x}\right)\right)^{-1/\theta_2}$
14	$\theta_0 x^{\theta_1} \exp(-\theta_2 x)$	$\vartheta_y = \theta_0(\theta_1/\theta_2) \exp(-\theta_1)$ $\vartheta_p = \theta_2^{\theta_1+1}$	$\theta_0 = \vartheta_y \left(\frac{1}{\vartheta_x}\right)^{\theta_1} \exp(\theta_1)$ $\theta_2 = \vartheta_p^{-1/(\theta_1+1)}$	$\vartheta_y \left(\frac{x}{\vartheta_x}\right)^{\theta_1} \exp\{\theta_1(1 - x/\vartheta_x)\}$ $\theta_1 : \theta_1 - \theta_x \vartheta_p^{-1/(\theta_1+1)}$

Tabela 3 (cont.) Reparametrizações desenvolvidas com ênfase na interpretação dos parâmetros de modelos de regressão não linear aplicados em Ciências Agrárias.

id	Modelo original	$\vartheta = g(\theta)$	$\theta_i = g^{-1}(\vartheta, \theta_{-i})$	Modelo reparametrizado
15	$\frac{\theta_a}{1 + \exp\{\theta_0 + \theta_1 x\}}$	$\vartheta_q = \frac{1}{\theta_1} \left(\log\left(\frac{1-q}{q}\right) - \theta_0 \right)$ $\vartheta_r = -\frac{4}{\theta_1}$	$\theta_0 = \log\left(\frac{1-q}{q}\right) - \theta_1 \vartheta_q$ $\theta_1 = -4\vartheta_r$	$\frac{\theta_a}{1 + \left(\frac{1-q}{q}\right) \exp\{-4\vartheta_r(x - \vartheta_q)\}}$
16	$\theta_a \exp\{-\exp\{\theta_0 + \theta_1 x\}\}$	$\vartheta_q = \frac{\log(-\log(q)) - \theta_0}{\theta_1}$	$\theta_1 = \frac{\vartheta_x}{\log(-\log(q)) - \theta_0}$	$\theta_a \exp\{\log(q) \exp\{\theta_0(1 - x/\vartheta_x)\}\}$
17	$\theta_r + \frac{\theta_s - \theta_r}{(1 + \exp\{\theta_a + x\}^{\theta_n})^{\theta_m}}$	$\vartheta_i = -\theta_a - \log(\theta_m)/\theta_n$ $\vartheta_s = -\frac{\theta_n(\theta_s - \theta_r)}{(1 - 1/\theta_m)^{\theta_m+1}}$	$\theta_a = -\vartheta_i - \log(\theta_m)/\theta_n$ $\theta_s - \theta_r = -\frac{\vartheta_s}{\theta_n} (1 + 1/\theta_m)^{\theta_m+1}$	$\theta_r - \frac{\vartheta_s}{\theta_n} \frac{(1 + 1/\theta_m)^{\theta_m+1}}{(1 + \exp\{\theta_n(x - \vartheta_i)\})/\theta_m)^{\theta_m}}$

3 ESTIMAÇÃO

Nessa seção será feita uma discussão sobre inferência sobre parâmetros em modelos de regressão não linear. A inferência baseada em verossimilhança será discutida, bem como inferência baseada na sua aproximação quadrática. Por fim, uma revisão do método delta será apresentada.

3.1 Verossimilhança

Seja $f(x, \theta)$ um modelo de regressão não linear considerado para descrever a média de uma variável aleatória Y . Considere que Y tenha distribuição normal com variância constante σ^2 . Resumidamente, podemos escrever esse modelo como

$$Y \sim \text{Normal}(\mu(x), \sigma^2)$$

$$\mu(x) = f(x, \theta).$$

A função de verossimilhança do modelo é dada por

$$L(\theta, \sigma^2) = \prod_{i=1}^n \phi(y_i, f(x_i, \theta), \sigma^2), \quad (3.1)$$

em que ϕ representa a função densidade da distribuição Normal. O estimador de máxima verossimilhança são os valores $(\hat{\theta}, \hat{\sigma}^2)$ que tornam máximo o valor de L . Para estimação, é conveniente trabalhar com o logaritmo da função de verossimilhança

$$\ell(\theta, \sigma^2) = \log L(\theta, \sigma^2). \quad (3.2)$$

Um intervalo de confiança para $\theta_i, i = 1, \dots, p$, baseado na função devi-

ance $D(\theta_i)$ é definido por $\theta_i : D(\theta_i) \leq c$ em que $D(\theta_i) = -2(\ell(\theta_i) - \ell(\hat{\theta}))$. A quantidade $\ell(\theta_i)$ representa o valor maximizado da log-verossimilhança com relação aos demais parâmetros para um valor fixo de θ_i enquanto que $\ell(\hat{\theta})$ é o valor máximo da log-verossimilhança com relação à todos os parâmetros. O valor c geralmente é definido de forma que o intervalo possua uma interpretação frequentista e se baseia na distribuição amostral da função deviance que é a distribuição qui-quadrado. Por exemplo, para obter um intervalo de confiança com cobertura nominal de 95% o valor de c é $\chi_1^2 = 3,84$.

Para encontrar os limites do intervalo de confiança para θ_i é necessário encontrar as raízes da equação $D(\theta_i) - c = 0$. No caso de modelos de regressão não linear, não é possível obter expressões fechadas para os limites do intervalo, então métodos numéricos são aplicados.

Os intervalos obtidos dessa forma apresentam alguma assimetria devido à verossimilhança ser uma função não simétrica em θ_i . O grau de assimetria diminui com o tamanho da amostra. O gráfico de $D(\theta_i)$ em função de θ_i , chamado de gráfico de perfil de log-verossimilhança para o parâmetro θ_i , permite visualizar tanto os limites do intervalo com relação ao valor estabelecido de c quanto o grau de simetria da função. Para Ritz e Streibig (2008), o gráfico de perfil é útil para acessar a extensão da curvatura do modelo na direção do parâmetro θ_i . Mais detalhes sobre o perfil de verossimilhança estão disponíveis em Bates e Watts (1988) e Venables e Ripley (2002).

Clarke (1987) aponta que a maior dificuldade no uso da função de verossimilhança para obter intervalos de confiança está na quantidade de trabalho envolvida para obter os seus limites. Tipicamente, isso requer sucessivas maximizações da verossimilhança com relação aos demais parâmetros para valores de θ_i numa sequência qualquer (MATTHEWS, 1988).

3.2 Aproximação quadrática da verossimilhança

Dada a restrição computacional para obtenção de intervalos baseados em verossimilhança, é comum simplificar a obtenção do intervalo de confiança para θ_i por meio de uma aproximação quadrática da função de log-verossimilhança ao redor de $\hat{\theta}$. Procedendo dessa forma temos

$$\begin{aligned} D(\theta_i) &= -2[\ell(\theta_i) - \ell(\hat{\theta}_i)] \\ &\approx -2 \left[\ell(\hat{\theta}_i) + (\theta_i - \hat{\theta}_i)\ell'(\hat{\theta}_i) + \frac{1}{2}(\theta_i - \hat{\theta}_i)^2\ell''(\hat{\theta}_i) - \ell(\hat{\theta}_i) \right]. \end{aligned}$$

Uma vez que $\hat{\theta}_i$ é estimador de máxima log-verossimilhança, $\ell'(\hat{\theta}_i) = 0$. Assim define-se o intervalo de confiança aproximado por meio de

$$D(\theta_i) = -(\theta_i - \hat{\theta}_i)^2\ell''(\hat{\theta}_i) \leq c, \quad (3.3)$$

com solução do tipo

$$\hat{\theta}_i \pm \sqrt{\frac{c}{\ell''(\hat{\theta}_i)}}. \quad (3.4)$$

Quando a distribuição considerada para a variável dependente é normal com variância constante, usar a aproximação da função de log-verossimilhança é equivalente a usar uma aproximação da função f para obter intervalos de confiança. Em modelos lineares, a soma de quadrado dos resíduos

$$\text{SQR}(\boldsymbol{\theta}) = \sum_{i=1}^n (y_i - f(x_i, \boldsymbol{\theta}))^2 \quad (3.5)$$

é quadrática e seus contornos de valor constante são elipsoides. Em modelos não lineares, se aproximarmos a função f por uma série de Taylor ao redor de $\hat{\theta}$, a soma de quadrados resultante da aproximação será quadrática. Isso resulta em intervalos individuais simétricos centrados em $\hat{\theta}_i$.

Segundo Seber e Wild (2003), fazendo uso de uma aproximação linear da função f e para uma amostra suficientemente grande, $\hat{\theta}$ apresentará distribuição amostral normal, ou seja,

$$\hat{\theta} \sim \text{Normal}(\theta, \sigma^2(\mathbf{F}^\top \mathbf{F})^{-1}), \quad (3.6)$$

em que $\mathbf{F} = \frac{\partial f(\mathbf{x}, \theta)}{\partial \theta^\top}$ representa a matriz de delineamento dada pela aproximação linear da função f avaliada em $\hat{\theta}$. Um intervalo de confiança de cobertura $1 - \alpha$ para θ_i é então obtido por

$$\hat{\theta}_i \pm t_{n-p} \sqrt{\hat{\sigma}_{ii}}, \quad (3.7)$$

em que t_{n-p} é o quantil $1 - \alpha/2$ da distribuição t com $n - p$ graus de liberdade, $\hat{\sigma}_{ii}$ representa a estimativa da variância de $\hat{\theta}_i$ e é o i -ésimo elemento da diagonal da matriz de variância-covariância $\hat{\sigma}^2(\mathbf{F}^\top \mathbf{F})^{-1}$, sendo $\hat{\sigma}^2 = \frac{\text{SQR}(\hat{\theta})}{n-p}$. Intervalos obtidos dessa forma, do tipo mais ou menos, são chamados de intervalos de Wald.

A vantagem do procedimento de Wald é que não requer a avaliação consecutiva de nenhuma função como é o intervalo baseado na verossimilhança, apenas do cálculo da matriz $\hat{\sigma}^2(\mathbf{F}^\top \mathbf{F})^{-1}$ após obter as estimativas dos parâmetros.

3.3 Método delta

O método delta é usado para aproximar a média e a variância de funções não lineares de variáveis aleatórias. Dentre suas aplicações, uma das mais comuns é relacionada à inferência sobre funções de parâmetros em modelos de regressão, como a razão entre parâmetros, transformação de um parâmetro, ou valor predito pelo modelo. Exemplos de funções de parâmetros estão na terceira coluna das tabelas 1 à 3.

Considere as variáveis aleatórias W_1, \dots, W_p , com $E(W_i) = \mu_i$, $\text{Var}(W_i) =$

$\sigma_i^2 = \sigma_{ii}$ e $\text{Cov}(W_i, W_j) = \sigma_{ij}$. Seja Σ_W a matriz de variância-covariância e $\boldsymbol{\mu}$ o vetor de médias para este vetor de variáveis aleatórias \mathbf{W} . Seja

$$T = g(W_1, \dots, W_p) = g(\mathbf{W}), \quad (3.8)$$

em que g é estritamente monótona e diferenciável em seu domínio (KHURI, 2003).

Fazendo a expansão em série de Taylor de g em torno dos valores esperados μ_1, \dots, μ_p , obtemos

$$T \approx g(\mu_1, \dots, \mu_p) + \sum_{i=1}^p g'_i(\mu_i)(W_i - \mu_i),$$

em que $g'_i(\mu_i) = \left. \frac{\partial g(W_1, \dots, W_p)}{\partial W_i} \right|_{W_i = \mu_i}$.

Consequentemente

$$E(T) \approx E(g(\mu_1, \dots, \mu_p)) + \sum_{i=1}^p E(g'_i(\mu_i)(W_i - \mu_i))$$

$$E(T) \approx g(\mu_1, \dots, \mu_p) = g(\boldsymbol{\mu}).$$

Da mesma forma, a variância de T baseada nessa aproximação é dada por

$$\text{Var}(T) \approx E((g(W_1, \dots, W_p) - g(\mu_1, \dots, \mu_p))^2)$$

$$\text{Var}(T) \approx E\left(\left(\sum_{i=1}^p g'_i(\mu_i)(W_i - \mu_i)\right)^2\right)$$

$$\text{Var}(T) \approx \sum_{i=1}^p g'_i(\mu_i)^2 \text{Var}(W_i) + \sum_{i < j} g'_i(\mu_i) g'_j(\mu_j) \text{Cov}(W_i, W_j)$$

$$\text{Var}(T) \approx \mathbf{D} \Sigma_W \mathbf{D}^\top, \quad \text{em que } \mathbf{D} = \left. \frac{\partial g(\mathbf{W})}{\partial \mathbf{W}^\top} \right|_{\mathbf{W} = \boldsymbol{\mu}}.$$

No contexto de regressão não linear, os estimadores $\hat{\theta}_1, \dots, \hat{\theta}_p$ são as variáveis aleatórias e $\hat{\vartheta} = g(\hat{\theta}_1, \dots, \hat{\theta}_p)$ é a função de interesse, ou parâmetro alvo de inferência. Dessa forma, o valor esperado e sua variância aproximados para $\hat{\vartheta}$ são calculados por

$$\begin{aligned} E(\hat{\vartheta}) &\approx g(\hat{\boldsymbol{\theta}}). \\ \text{Var}(\hat{\vartheta}) &\approx \mathbf{D}\boldsymbol{\Sigma}_\theta\mathbf{D}^\top, \quad \text{em que } \mathbf{D} = \left. \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \end{aligned}$$

Nas aplicações, no entanto, a matriz de variância-covariância estimada $\hat{\boldsymbol{\Sigma}}_\theta$ é usada no lugar de $\boldsymbol{\Sigma}_\theta$.

Se $\hat{\vartheta}$ tem distribuição amostral Normal, um intervalo de confiança de cobertura nominal $1 - \alpha$ para ϑ é obtido por

$$E(\hat{\vartheta}) \pm z_{1-\alpha/2} \sqrt{\text{Var}(\hat{\vartheta})}, \quad (3.9)$$

em que $z_{1-\alpha/2}$ é o quantil $1 - \alpha/2$ da distribuição Normal padrão.

4 CONSIDERAÇÕES FINAIS

Os métodos descritos são os meios disponíveis pelos quais pode-se fazer inferência sobre parâmetros. A vantagem da abordagem de verossimilhança está na obtenção dos intervalos de confiança que são menos sensíveis à medidas de não linearidade e suposições assintóticas, porém, tem a desvantagem do elevado custo computacional. O uso da aproximação quadrática ou abordagem de Wald, por outro lado, têm baixo custo computacional mas faz uso de argumentos assintóticos que podem não ser completamente verificados. Quando o modelo de regressão é não linear, testes de hipóteses ou intervalos de confiança baseados na função de log-verossimilhança e na sua aproximação quadrática (Wald) podem levar à conclusões diferentes (GALLANT, 1987; DRAPER; SMITH, 1998). Diversos autores apontam que a extensão dessas diferenças são dependentes da curvatura do modelo de regressão não linear que por sua vez é sensível a parametrização (BEALE, 1960; BATES; WATTS, 1980; HOUGAARD, 1982; RATKOWSKY, 1983; RATKOWSKY, 1986; CLARKE, 1987; MATTHEWS, 1988; HOUGAARD, 1988; BARROZO et al., 2004; SARADA, 2005; VILA; GAUCHI, 2007; ROSS; SARADA, 2010; ZEVIANI et al., 2012)

Em ambas as abordagens, o modelo de regressão não linear pode ser avaliado sob qualquer parametrização, de forma que é possível escolher àquela que apresente o melhor compromisso entre propriedades estatísticas e de interpretação. Além disso, por ambas as abordagens, o modelo de regressão pode ser estendido para acomodar o efeito de outras variáveis independentes, como efeito fixo de fatores categóricos e contínuos e o efeito aleatório associado às unidades de amostragem (PINHEIRO; BATES, 2000). A capacidade de modelar os parâmetros só está disponível para aqueles parâmetros do modelo e que, por meio de reparamen-

trização, os parâmetros de interesse podem ser representados. Estando o modelo em uma parametrização interpretável, fica simples incorporar informação prévia por meio de prioris ao considerar inferência bayesiana (DENISON et al., 2002).

O método delta, diferente dos anteriores, só atende ao objetivo de fazer inferência para funções dos parâmetros após estimação dos parâmetros do modelo. Ou seja, o método delta não permite modelar o parâmetro nem incorporar prioris. No caso de g ser uma função linear e ter distribuição normal multivariada, como é o caso nos modelo de regressão linear múltipla, $\hat{\theta}$ teria distribuição normal por ser uma função linear de variáveis aleatórias normais (ALPER; GELB, 1990). No entanto, no caso dos modelos não lineares, $\hat{\theta}$ têm distribuição amostral normal apenas para amostras grandes e para amostras pequenas a distribuição amostral depende da curvatura do modelo (BEALE, 1960; BATES; WATTS, 1980; RATKOWSKY, 1986; CLARKE, 1987; MAZUCHELI; ACHCAR, 2002). Além do mais, a qualidade da aproximação depende do quão linear g é ao redor de $\hat{\theta}$ (melhor quanto mais linear) e do tamanho das variâncias de cada θ (melhor quanto menor a variância). Dessa forma, concluí-se que $\hat{\theta}$ só apresentará distribuição normal em pequenas amostras se g for uma função linear e $\hat{\theta}$ for normal multivariada, o que só acontece nos modelos de regressão linear múltipla. No entanto, para Bender (1996), esse método produz resultados úteis quando a amostra é suficientemente grande (> 100 observações) onde a suposição de normalidade para $\hat{\theta}$ não chega ser restritiva pois os métodos de estimação frequentemente produzem estimadores com distribuição assintoticamente normal, como o caso dos estimadores de máxima verossimilhança.

A conclusão que se antecipa é que, uma vez que é possível reparametrizar o modelo para o parâmetro de interesse, inferência baseada na verossimilhança deve ser considerada, em segundo, sua aproximação quadrática, visto a capacidade

de modelagem permitida por tais abordagens. Além do mais, as parametrizações devem ser avaliadas, seja por meio de medidas de curvatura, gráficos de perfil ou simulação, e deve ser escolhida àquela que tenha melhor compromisso entre propriedades estatísticas e de interpretação.

REFERÊNCIAS

ALPER, J. S.; GELB, R. I. Standard errors and confidence intervals in nonlinear regression: comparison of Monte Carlo and parametric statistics. **The Journal of Physical Chemistry**, Washington, v. 94, n. 11, p. 4747–4751, May 1990.

ATKINSON, A. C. et al. Optimum experimental designs for properties of a compartmental model. **Biometrics**, Washington, v. 49, n. 2, p. 325–337, Dec. 1993.

BARROZO, M.; SARTORI, D.; FREIRE, J. A study of the statistical discrimination of the drying kinetics equations. **Food and Bioprocess Technology**, Dorchester, v. 82, n. 3, p. 219–225, Sept. 2004.

BAS, D.; BOYACI, I. H. Modeling and optimization I: usability of response surface methodology. **Journal of Food Engineering**, v. 78, n. 3, p. 836–845, Feb. 2007.

BATES, D.; WATTS, D. Relative curvature measures of nonlinearity. **Journal of the Royal Statistical Society. Series B (Methodological)**, London, v. 42, n. 1, p. 1–25, Jan. 1980.

BATES, D. M.; WATTS, D. G. **Nonlinear Regression Analysis and Its Applications**. New York: John Wiley & Sons, 1988. 365 p.

BEALE, E. M. Confidence regions in non-linear estimation. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 22, n. 1, p. 41–76, Jan. 1960.

BECKER, P. M. et al. Validation of growth as measurand for bacterial adhesion to food and feed ingredients. **Journal of Applied Microbiology**, Oxford, v. 103, n. 6, p. 2686–96, Dec. 2007.

BENDER, R. Calculating confidence intervals for summary measures of individual curves via nonlinear regression models. **International Journal of Bio-Medical Computing**, Limerick, v. 41, n. 1, p. 13–18, Mar. 1996.

BLEASDALE, J. K. A.; NELDER, J. A. Plant population and crop yield. **Nature**, London, v. 188, n. 4747, p. 342–342, Oct. 1960.

BOCK, B. R.; SIKORA, F. J. Modified-quadratic/plateau model for describing plant responses to fertilizer. **Soil Science Society of America Journal**, Madison, v. 54, n. 6, p. 1784, Nov. 1990.

BONAT, W. H. et al. **Métodos computacionais para inferência com aplicações em R**. 1. ed. João Pessoa: Associação Brasileira de Estatística, 2012. 260 p.

BOX, G.; LUCAS, H. Design of experiments in non-linear situations. **Biometrika**, London, v. 46, p. 77–90, June 1959.

BULLOCK, D. G.; BULLOCK, D. S. Quadratic and quadratic-plus-plateau models for predicting optimal nitrogen rate of corn: a comparison. **Agronomy Journal**, Madison, v. 86, n. 1, p. 191, Jan. 1994.

CARNEIRO, F. F. et al. Minimum number of common bean plants per plot to assess field resistance to white mold. **Crop Breeding and Applied Biotechnology**, v. 11, n. 4, p. 358–364, Dec. 2011.

CHENG, R.; EVANS, B.; ILES, T. Embedded models in non-linear regression. **Journal of the Royal Statistical Society**. Series B (Methodological), London, v. 54, n. 3, p. 877–888, Aug. 1992.

CLARKE, G. Approximate confidence limits for a parameter function in nonlinear regression. **Journal of the American Statistical Association**, New York, v. 82, n. 397, p. 221–230, Mar. 1987.

CONGLETON, W.; EVERETT, R. Application of the incomplete gamma function to predict cumulative milk production. **Journal of Dairy Science**, v. 63, n. 1, p. 109–119, Jan. 1980.

DAMGAARD, C.; BORKSTED, B. Transgenic insect resistant Arabidopsis may show chaotic population dynamic. **Ecological Complexity**, v. 1, n. 3, p. 261–265, Sept. 2004.

DENISON, D. G. T. et al. **Bayesian Methods for Nonlinear Classification and Regression**. 1. ed. [S.l.]: Wiley, 2002. 296 p.

DEXTER, A. Soil physical quality: Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. **Geoderma: an international journal of soil science**, Amsterdam, v. 120, n. 3-4, p. 201–214, June 2004.

DHANOA, M. S. A note on an alternative form of the lactation model of Wood. **Animal Production**, Bletchley, v. 32, n. 03, p. 349–351, June 1981.

DRAPER, N.; SMITH, H. **Applied regression analysis**. 3. ed. New York: Wiley-Interscience, 1998. 706 p.

- FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, Abingdon, v. 31, n. 7, p. 799–815, Aug. 2004.
- FORD, I.; TITTERINGTON, D. M.; KITSOS, C. P. Recent advances in nonlinear experimental design. **Technometrics**, New York, v. 31, n. 1, p. 49–60, Mar. 1989.
- FRANCE, J.; THORNLEY, J. **Mathematical models in agriculture**. 2. ed. [S.l.]: Cabi, 2006. 848 p.
- GALLANT, A. R. **Nonlinear statistical models**. 1. ed. [S.l.]: Wiley, 1987. 610 p.
- GENUCHTEN, M. T. van. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. **Soil Science Society of America Journal**, Madison, v. 44, n. 5, p. 892–898, May 1980.
- GOUTELLE, S. et al. The Hill equation: a review of its capabilities in pharmacological modelling. **Fundamental & Clinical Pharmacology**, Paris, v. 22, n. 6, p. 633–48, Dec. 2008.
- GROOT, J. C. J. et al. Multiphasic analysis of gas production kinetics for in vitro fermentation of ruminant feeds. **Animal Feed Science and Technology**, Amsterdam, v. 64, n. 1, p. 77–89, Dec. 1996.
- GUIMARÃES, V. P. et al. Utilização de funções matemáticas no estudo da curva de lactação em caprinos. **Revista Brasileira de Zootecnia**, Viçosa, v. 35, n. 2, p. 535–543, Apr. 2006.
- HILL, A. V. The combinations of haemoglobin with oxygen and with carbon monoxide. **Biochemical Journal**, London, v. 7, n. 5, p. 471–480, Oct. 1913.
- HOUGAARD, P. Parameterizations of non-linear models. **Journal of the Royal Statistical Society. Series B (Methodological)**, London, v. 44, n. 2, p. 244–252, Aug. 1982.
- HOUGAARD, P. The asymptotic distribution of nonlinear regression parameter estimates: improving the approximation. **International Statistical Review**, Edinburg, v. 56, n. 3, p. 221–233, Dec. 1988.
- KHURI, A. I. **Advanced calculus with applications in statistics**. 2. ed. Hoboken: J. Wiley, 2003. 673 p.
- KHURI, A. I. **Response surface methodology and related topics**. Singapore: World Scientific, 2006. 457 p.

- KRANZ, J. (Ed.). **Epidemics of plant diseases**. 1. ed. Berlin: Springer, 1990. 268 p.
- LYLES, R. H. et al. Nonlinear model-based estimates of IC(50) for studies involving continuous therapeutic dose-response data. **Contemporary Clinical Trials**, London, v. 29, n. 6, p. 878–86, Nov. 2008.
- MATTHEWS, D. Likelihood-based confidence intervals for functions of many parameters. **Biometrika**, London, v. 75, n. 1, p. 139–144, Sept. 1988.
- MAZUCHELLI, J.; ACHCAR, J. A. Algumas considerações em regressão não linear. **Acta Scientiarum: technology**, Maringá, v. 24, n. 6, p. 1761–1770, Nov. 2002.
- MESHKAT, N.; ANDERSON, C.; DISTEFANO, J. J. Finding identifiable parameter combinations in nonlinear ODE models and the rational reparameterization of their input-output equations. **Mathematical biosciences**, New York, v. 233, n. 1, p. 19–31, Sept. 2011.
- MICHAELIS, L.; MENTEN, M. Die kinetik der invertinwirkung. **Biochemische zeitung**, v. 49, p. 333–369, Feb. 1913.
- MUGGEO, V. M. R. Estimating regression models with unknown break-points. **Statistics in Medicine**, New York, v. 22, n. 19, p. 3055–3071, Oct. 2003.
- OLORI, V. et al. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. **Livestock Production Science**, Amsterdam, v. 58, n. 1, p. 55–63, Mar. 1999.
- OMUTO, T. et al. Nonlinear mixed effect modelling for improved estimation of water retention and infiltration parameters. **Journal of Hydrology**, Amsterdam, v. 330, n. 3-4, p. 748–758, Nov. 2006.
- PÁZMAN, A.; PRONZATO, L. Nonlinear experimental design based on the distribution of estimators. **Journal of Statistical Planning and Inference**, Amsterdam, v. 33, n. 3, p. 385–402, Dec. 1992.
- PEEK, M. et al. Physiological response curve analysis using nonlinear mixed models. **Oecologia**, Berlin, v. 132, n. 2, p. 175–180, July 2002.
- PEIXOTO, A. P. B.; FARIA, G. A.; MORAIS, A. R. Modelos de regressão com platô na estimativa do tamanho de parcelas em experimento de conservação in vitro de maracujazeiro. **Ciência Rural**, Santa Maria, v. 41, n. 11, p. 1907–1913, Nov. 2011.

- PINHEIRO, J.; BATES, D. **Mixed-effects models in S and S-plus**. 1. ed. New York: Springer, 2000. 548 p.
- RATKOWSKY, D. A. **Nonlinear regression modeling**. New York: M. Dekker, 1983. 276 p.
- RATKOWSKY, D. A. Statistical properties of alternative parameterizations of the von Bertalanffy growth curve. **Canadian Journal of Fisheries and Aquatic Sciences**, Ottawa, v. 43, n. 4, p. 742–747, Apr. 1986.
- RATKOWSKY, D. A. **Handbook of nonlinear regression models**. New York: M. Dekker, 1990.
- REKAYA, R.; WEIGEL, K.; GIANOLA, D. Hierarchical nonlinear model for persistency of milk yield in the first three lactations of Holsteins. **Livestock Production Science**, Amsterdam, v. 68, n. 2-3, p. 181–187, Mar. 2001.
- RITZ, C.; STREIBIG, J. C. **Nonlinear regression with R**. 1. ed. New York: Springer, 2008. 142 p.
- ROSS, G. J.; SARADA, C. Reparameterization of nonlinear statistical models: a case study. **Journal of Applied Statistics**, Abingdon, v. 37, n. 12, p. 2015–2026, Dec. 2010.
- SARADA, C. On appropriate reparameterization of a nonlinear statistical model. **Journal of the Indian Society of Agricultural Statistics**, New Delhi, v. 59, n. 3, p. 237–242, Nov. 2005.
- SCHABENBERGER, O.; PIERCE, F. J. **Contemporary statistical models for the plant and soil Sciences**. 1. ed. Boca Raton: CRC, 2002. 738 p.
- SCHABENBERGER, O. et al. Statistical tests for hormesis and effective dosages in herbicide dose response. **Agronomy Journal**, Madison, v. 91, n. 4, p. 713–721, Mar. 1999.
- SEBER, G. A. F.; WILD, C. J. **Nonlinear regression**. Hoboken: Wiley, 2003. 752 p.
- SEGARRA, J.; JEGER, M. J.; BOSCH, F. van den. Epidemic dynamics and patterns of plant diseases. **Phytopathology**, Saint Paul, v. 91, n. 10, p. 1001–10, Oct. 2001.
- SERAFIM, M. E. et al. Umidade do solo e doses de potássio na cultura da soja. **Revista Ciência Agronômica**, Fortaleza, v. 43, n. 2, p. 222–227, July 2012.

- SHUAI, X.; ZHOU, Z.; YOST, R. S. Using segmented regression models to fit soil nutrient and soybean grain yield changes due to liming. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 8, n. 2, p. 240–252, June 2003.
- TJORVE, E. Shapes and functions of species-area curves: a review of possible models. **Journal of Biogeography**, Blackwell, v. 30, n. 6, p. 827–835, May 2003.
- TSAI, C. Power transformations and reparameterizations in nonlinear regression models. **Technometrics**, New York, v. 30, n. 4, p. 441–448, Mar. 1988.
- VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S**. 4ed. ed. New York: Springer, 2002. 495 p.
- VIETH, E. Fitting piecewise linear regression functions to biological responses. **Journal of Applied Physiology**, Bethesda, v. 67, n. 1, p. 390–396, July. 1989.
- VILA, J.-P.; GAUCHI, J.-P. Optimal designs based on exact confidence regions for parameter estimation of a nonlinear regression model. **Journal of Statistical Planning and Inference**, Amsterdam, v. 137, n. 9, p. 2935–2953, Sept. 2007.
- WATKINSON, A. Density-dependence in single-species populations of plants. **Journal of Theoretical Biology**, London, v. 83, n. 2, p. 345–357, Mar. 1980.
- WOOD, P. D. P. Algebraic model of the lactation curve in cattle. **Nature**, London, v. 216, n. 5111, p. 164–165, Oct. 1967.
- YIN, X. et al. A nonlinear model for crop development as a function of temperature. **Agricultural and Forest Meteorology**, Amsterdam, v. 77, n. 1-2, p. 1–16, Nov. 1995.
- ZEVIANI, W. M. et al. Modelos não lineares para a liberação de potássio de esterco animal em latossolos. **Ciência Rural**, Santa Maria, v. 42, n. 10, p. 1789–1796, Oct. 2012.

CAPÍTULO 2 MODELO NÃO LINEAR PARA O NÍVEL DE DANO ECONÔMICO DA DESFOLHA NO ALGODOEIRO

RESUMO

O efeito da desfolha sobre a qualidade e produtividade das culturas é informação fundamental para definir estratégias de manejo, como intensidade e frequência de pastejo e colheita até o estabelecimento de níveis de dano econômico de forma a auxiliar decisões sobre o controle de pragas desfolhadoras. Para a cultura do algodão, assim como para outras tantas, a redução da produção pela desfolha pode ser representada por uma função não linear monótona não crescente. Diversos modelos podem satisfazer essa restrição, no entanto, existe a preocupação de inferir sobre o nível de dano econômico, ϑ_q , pelo ajuste de um modelo. Dados de produção-desfolha do algodoeiro em função do estágio fenológico são considerados para inferir sobre o nível de dano econômico com os seguintes objetivos: 1) propor uma parametrização de modelo que representasse o parâmetro, 2) avaliar parametrizações alternativas por meio de medidas de não linearidade, 3) aplicar inferência baseada em verossimilhança, 4) selecionar um modelo para descrever a relação entre produção e desfolha do algodoeiro em função do estágio fenológico. O modelo reparametrizado apresentou menores medidas de não linearidade nos estágios fenológicos com pronunciada relação não linear. Nos restantes, as medidas de curvatura, as correlações dos estimadores e os gráficos de perfil de verossimilhança indicaram que um sub-modelo deveria ser considerado.

Palavras-chave: Interpretação de parâmetros. Verossimilhança. Método delta. Curvatura. *Gossypium hirsutum*.

ABSTRACT

The effect of defoliation on the quality and productivity of crops is essential information to define management strategies, such as intensity and frequency of grazing and harvesting and the establishment of economic threshold in order to aid decisions about controlling defoliating pests. For the cotton, as well as many others crops, reduction of production by defoliation can be represented by a nonincreasing monotone function. Several models can satisfy this restriction, however, there is concern about inferring the economic damage level, ϑ_q , by adjusting a model. Yield-defoliation data of cotton due to the phenological stage are considered to infer about the economic damage level with the following objectives: 1) to propose a model that represents the parameter ϑ_q , 2) evaluate alternative parameterizations through measures of nonlinearity, 3) apply inference based on likelihood, 4) select a model to describe the relationship between yield and defoliation of cotton in each phenological stage. The reparametrized model had lower measures of nonlinearity in phenological stages with pronounced nonlinear relationship. In the others, the measures of curvature, the correlations of the estimators and likelihood profile plots indicated that a sub-model should be considered.

Key-words: Parameter interpretation. Likelihood. Delta method. Curvature. *Gossypium hirsutum*.

1 INTRODUÇÃO

Em condições de campo, as culturas estão sujeitas à perdas de área foliar por diferentes causas, dentre elas o pastejo e a colheita periódica das folhas, o ataque de insetos desfolhadores e de doenças que causam sua queda ou necrose, as chuvas de granizo e a própria senescência natural são as mais frequentes. Uma desfolha significativa reduz o potencial fotossintético e, dependendo da intensidade e fase de crescimento da planta, ocasiona prejuízos à produção (PAINTER; DETLING, 1981; KLUBERTANZ et al., 1996). Algumas doenças e pragas, fitotoxicidade de pesticidas ou adubos, granizo e certas injúrias mecânicas são eventos comuns que causam desfolha em áreas de cultivo de algodão e que podem prejudicar a produção e qualidade do produto dessa cultura (SILVA et al., 2012).

A avaliação dos efeitos da desfolha são fundamentais para estabelecer estratégias de manejo, como uso de adubação e irrigação de recuperação. Para a bovinocultura, por exemplo, tais informações permitem definir a periodicidade e intensidade de pastejo ou corte da forragem a fim de manter a sua qualidade e produtividade (EDELSTEN; CORRALL, 1979; MOTAZEDIAN; SHARROW, 1990; ZEWDU, 2003; LAUER et al., 2004; VOLESKY et al., 2011). Tais informações são igualmente relevantes para outras culturas de importância econômica, como a soja (KAWAMOTO et al., 1986; KLUBERTANZ et al., 1996), o arroz (BERTONCELLO et al., 2011), o feijão (SCHAAFSMA; ABLETT, 1994), o girassol (LIMA JUNIOR et al., 2010), o milho (ZEWDU, 2003; LAUER et al., 2004), a batata (HARE, 1980), o eucalipto (BATTAGLIA et al., 2011) e o algodão (LUDWIG et al., 1965; GUTIERREZ et al., 1975; FAIRCLOTH et al., 2004a; FAIRCLOTH et al., 2004b)

Muitos estudos visam relacionar o efeito da desfolha com outras variáveis

do sistema produtivo como a fase de crescimento em que ocorre (SILVA et al., 2012), o período do ano (EDELSTEN; CORRALL, 1979; HARE, 1980), a adubação (COUGHENOUR et al., 1990; WACHENDORF et al., 2004) e a densidade de pragas (ZEHNDER; EVANYLO, 1988). A aplicação de níveis de desfolha artificial permite determinar a intensidade que a cultura pode suportar, inclusive em associação à outros fatores, como estágio fenológico, e seu impacto nos componentes de produção e qualidade dos produtos finais. Com isso pode-se estabelecer o nível de dano econômico que é fundamental para adoção de estratégias de manejo economicamente viáveis. Além disso, poupar uma aplicação de insumo para controle da praga não é relevante apenas no âmbito econômico da atividade agrícola mas também reduz os impactos ecológicos decorrentes da aplicação, como contaminação do solo e água, e o extermínio de inimigos naturais. Kulman (1971) reconhece que a desfolha em culturas perenes diminui a taxa de crescimento, a qualidade da madeira e facilita o ataque de organismos secundários, como doenças e plantas invasoras, mas que estratégias de controle devem considerar o custo do controle diante do prejuízo previsto.

Em condições normais de manejo da cultura do algodão, com a densidade populacional adequada, a produção é uma função monótona não crescente do nível de desfolha e depende do estágio fenológico em que ocorre (FAIRCLOTH et al., 2004a; SILVA et al., 2012). Em situações como essa, em que existe uma informação prévia sobre a relação entre as variáveis, os modelos de regressão não linear podem ser admitidos pois permitem, a partir dessas informações, escolher, elaborar ou reparametrizar um modelo que descreva apropriadamente essa relação e possua parâmetros de significado prático, o que facilita a interpretação dos resultados (SCHABENBERGER; PIERCE, 2002).

O objetivo com este trabalho é desenvolver e parametrizar modelos de

regressão não linear para estimar o nível de dano econômico da desfolha na produção algodoeiro. O enfoque, além do procedimento de elaboração dos modelos, será sobre a avaliação dos métodos de inferência disponíveis para parâmetros.

2 MODELO

A relação monótona não crescente entre produção e desfolha é a informação preliminar considerada para elaborar um modelo. Dentre as funções matemáticas que atendem à essa imposição, tem-se o modelo potência, chamado de modelo Herschel-Bulkley por Cheng et al. (1992), como opção,

$$f(x) = \theta_0 - \theta_1 x^{\theta_2}, \quad x \geq 0. \quad (2.1)$$

A desfolha, x (adimensional), assume valores entre 0 e 1. A produção normal, com unidade de medida representada por Y , prevista sem haver desfolha é representada pelo parâmetro θ_0 (Y), ou seja, $f(0) = \theta_0$. A redução na produção normal ao ocorrer uma desfolha total é $\theta_1 \geq 0$, ou seja, $f(0) - f(1) = \theta_1$. O parâmetro adimensional $\theta_2 > 0$ é um parâmetro de forma dessa relação, que é côncava se $\theta_2 > 1$, convexa se $0 < \theta_2 < 1$ e linear se $\theta_2 = 1$. É conveniente reescrever o modelo considerando a transformação $\theta_2 = \exp\{\theta_c\}$ uma vez que a função exponencial é positiva e que isso não compromete a interpretação do modelo em θ_2 que é apenas um parâmetro de forma. Dessa forma,

$$f(x) = \theta_0 - \theta_1 x^{\exp\{\theta_c\}}, \quad x \geq 0, \quad (2.2)$$

é uma função côncava para $\theta_c > 0$, convexa para $\theta_c < 0$ e linear quando $\theta_c = 0$ (Figura 1).

O parâmetro de interesse é o nível de dano econômico, ϑ_q , da desfolha sobre a produção da cultura. Tal quantidade é definida como o valor de desfolha para o qual a produção diminui $q > 0$ unidades com relação a produção normal θ_0 , ou seja, $q = f(0) - f(\vartheta_q)$ (Figura 1). Uma diminuição superior à q corresponde à

prejuízo econômico. O valor comercial do que é produzido e os custos de controle da desfolha, como insumos e aplicação, são necessários para estabelecer o valor de q , que por sua vez depende do ano agrícola.

Uma vez definido o valor de q , é possível estimar ϑ_q a partir de dados de produção-desfolha ao considerar o ajuste do modelo em (2.2). Com valores para os parâmetros de (2.2), ϑ_q é calculado por

$$\theta_0 - q = \theta_0 - \theta_1 \vartheta_q^{\exp(\theta_c)}, \text{ resultanto em} \quad (2.3)$$

$$\vartheta_q = \left(\frac{q}{\theta_1} \right)^{\exp(-\theta_c)}.$$

Com essa expressão pode-se obter uma estimativa pontual para ϑ_q após ajuste de (2.2) aos dados. No entanto, como nas situações reais não se conhecem os valores dos parâmetros, o valor de ϑ_q será função das estimativas dos parâmetros em (2.2) e por isso é necessário quantificar a incerteza sobre sua estimativa pontual. O método delta é usualmente aplicado nessas situações para obter um erro-padrão para estimativa e a partir disso construir um intervalo de confiança.

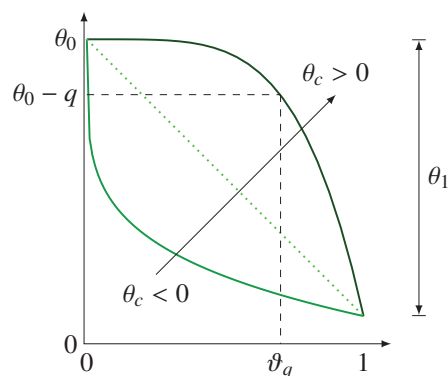


Figura 1 Modelo potência com destaque para o significado gráfico dos parâmetros e comportamento com relação ao parâmetro θ_c

Uma forma completamente direta de fazer inferência sobre ϑ_q é incorporá-lo ao modelo por meio de uma reparametrização. Para chegar à reparametrização do modelo, aplica-se o seguinte procedimento de três etapas:

1. Expressar o parâmetro de interesse como função dos elementos de θ , ou seja, $\vartheta = g(\theta)$;
2. Escolher um dos elementos de $\theta = (\theta_i, \theta_{-i})$ para ser colocado em função de ϑ de tal forma a obter $\theta_i = h(\theta_{-i}, \vartheta)$;
3. Substituir θ_i no modelo pela expressão obtida no passo anterior, $h(\theta_{-i}, \vartheta)$, fazendo as simplificações convenientes.

A primeira etapa do procedimento corresponde a definição de ϑ_q em (2.3). Sendo ϑ_q função dos parâmetros θ_1 e θ_c temos duas alternativas mas como θ_1 possui interpretação é vantajoso em termos de interpretação substituir θ_c por ϑ_q . Sendo assim, para a segunda etapa tem-se

$$\exp\{\theta_c\} = \frac{\log(q) - \log(\theta_1)}{\log(\vartheta_q)}. \quad (2.4)$$

Com isso, substituí-se (2.4) em (2.2) e obtém-se

$$f(x) = \theta_0 - \theta_1 x^{\frac{\log(q) - \log(\theta_1)}{\log(\vartheta_q)}}. \quad (2.5)$$

Ao considerar uma parametrização alternativa, deve-se verificar os espaços paramétricos dos parâmetros representados e os sub-modelos gerados por restrições paramétricas, equivalentes às existentes para o modelo original. O espaço paramétrico corresponde à ϑ_q é $0 < \vartheta_q < 1$. Além disso se que $\log(q) - \log(\theta_1) < \log(\vartheta_q)$ a função é convexa, para $\log(q) - \log(\theta_1) > \log(\vartheta_q)$ é concava e para $\log(q) - \log(\theta_1) = \log(\vartheta_q)$, que é equivalente à $q/\theta_1 = \vartheta_q$, a função é linear. Dessa

forma, testar a adequação do modelo com três parâmetros contra o modelo de dois parâmetros

$$f(x) = \theta_0 - \theta_1 x, \quad (2.6)$$

nada mais é que um teste sobre a concavidade nula ou relação linear entre produção e desfolha. No entanto, como o objetivo é sobre o nível de dano econômico, é apropriado considerar a parametrização

$$f(x) = \theta_0 - \frac{q}{\vartheta_q} x, \quad (2.7)$$

pois se $\theta_c = 0$ então $\theta_1 = \frac{q}{\vartheta_q}$. Além dessas restrições que induzem o sub-modelo linear simples, para esse modelo em particular, temos que sub-modelos são gerados para valores do expoente de x tendendo aos extremos da reta real positiva, 0 e ∞ ,

$$\lim_{\exp\{\theta_c\} \rightarrow \infty} = \lim_{\vartheta_q \rightarrow 1} = \begin{cases} \theta_0 & , 0 \leq x < 1 \\ \theta_0 - \theta_1 & , x = 1 \end{cases},$$

$$\lim_{\exp\{\theta_c\} \rightarrow 0} = \lim_{\vartheta_q \rightarrow 0} = \begin{cases} \theta_0 - \theta_1 & , x > 0 \\ \theta_0 & , x = 0 \end{cases}.$$

Nos limites o modelo retorna um valor constante, θ_0 ou $\theta_0 - \theta_1$, quando $\vartheta_q \rightarrow 1$ ou $\vartheta_q \rightarrow 0$. Em outras palavras, tais restrições reduzem o modelo de três parâmetros ao modelo com apenas o intercepto. No entanto, o mesmo acontece quando $\theta_1 \rightarrow 0$ e dessa forma, tem-se o que se conhece por falta de identificabilidade na borda do espaço paramétrico que é, nesse caso, também um indicador de não haver efeito de desfolha sobre a produção.

3 MATERIAL E MÉTODOS

Os dados considerados para ajuste do modelo são de um experimento, em casa de vegetação, com a cultura do algodão (*Gossypium hirsutum*). As unidades experimentais foram 2 plantas por vaso para registro da produção total de pluma com caroço (g). Os fatores estudados foram o nível de desfolha artificial (0, 25, 50, 75 e 100%), feita com tesoura em cada uma das folhas da planta conforme tais níveis, combinados com o estágio fenológico no qual a desfolha foi realizada (vegetativo, presença de botão floral, florescimento, presença de maçã e presença de capulho). O delineamento completamente ao acaso foi utilizado com cinco repetições, perfazendo $5 \times 5 \times 5 = 125$ unidades experimentais. O experimento foi realizado nas dependências da Universidade Federal de Grande Dourados no ano agrícola de 2007. Mais informações disponíveis em Silva et al. (2012). Na Figura 2 tem-se o diagrama de dispersão dos valores observados de peso de capulhos

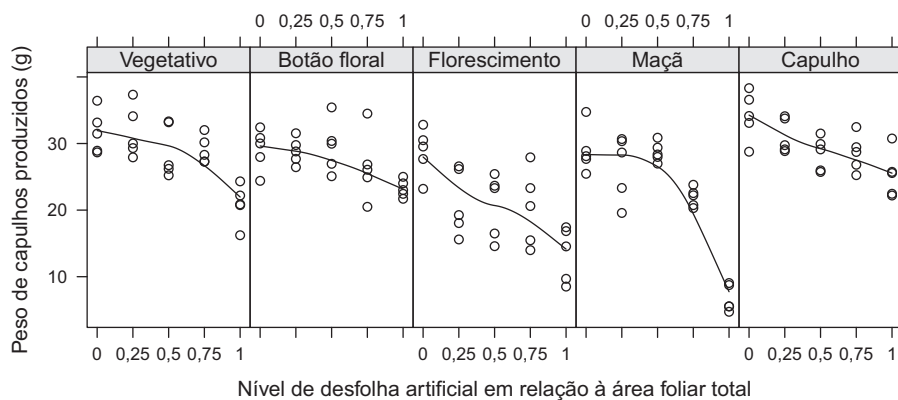


Figura 2 Peso de capulhos produzidos (g) em cada estágio fenológico como função dos níveis de desfolha artificial. Curvas suaves entre os pontos representam as tendências centrais

produzidos (g) em cada estágio fenológico como função dos níveis de desfolha.

Os modelos (2.2) e (2.5) foram ajustadas aos dados separado por estágio fenológico. Considerou-se $q = 5$ g como uma queda na produção correspondente à dano econômico. Tal proposta, ainda que empírica, baseou-se no fato que 5 g perfazem uma redução próxima a 17% na produção sem desfolha que foi ao redor de 30 g (Figura 2).

As medidas de curvatura de Bates e Watts (1980) foram calculadas para avaliar as propriedades inferenciais das parametrizações. Inferência para o parâmetro nível de dano econômico ϑ_q foram obtidas por três formas: 1) por meio do modelo original seguido da aplicação do método delta, 2) por meio do modelo reparametrizado baseado na distribuição assintótica dos estimadores e 3) por meio do modelo reparametrizado baseado em inferência por verossimilhança.

4 RESULTADOS E DISCUSSÃO

Os ajustes dos modelos aos dados convergiram para os cinco estágios fenológicos, considerando o máximo de 50 interações. Valores iniciais baseados na inspeção do diagrama de dispersão foram considerados (Figura 2). As estimativas dos parâmetros comuns, θ_0 e θ_1 , foram idênticas, em termos pontuais e intervalares, nas duas parametrizações, como de fato devem ser pois ambas parametrizações descrevem o mesmo modelo (Tabela 1). Percebeu-se um resultado alarmante para o estágio de florescimento, no qual os intervalos de confiança para θ_0 e θ_1 foram demasiado amplos, superando inclusive a amplitude média de variação dos dados, de aproximadamente 5 g para cima e para baixo, ao redor da curva de tendência.

As estimativas de θ_c foram positivas nos estágios vegetativo, botão floral e maçã o que indica uma função côncava, e negativas nos demais estágios o que indica uma função convexa. No entanto, os intervalos de confiança para θ_c no botão floral, florescimento e capulho contiveram o valor zero que indica que o sub-modelo de dois parâmetros, o linear simples, pode ser uma descrição adequada da relação produção-desfolha tanto quanto o modelo potência.

As estimativas pontuais e intervalares para ϑ_q pelo modelo reparametrizado foram idênticas as obtidas pelo método delta (Tabela 1). Tal equivalência decorre do fato de intervalos de confiança assintóticos serem obtidos por aproximação linear do modelo não linear ao obter a matriz de covariância assintótica dos parâmetros da mesma forma como faz o método delta ao aproximar linearmente (2.3) para obter o erro-padrão assintótico de ϑ_q (COX, 1990). Embora isso possa indicar não haver vantagem no uso do modelo reparametrizado, reconhece-se que intervalos de confiança assintóticos e o método delta fazem uso de argumentos assintóticos (BEALE, 1960; ALPER; GELB, 1990; BENDER, 1996) e que inter-

valos baseados na verossimilhança são mais adequados para representar a incerteza sobre os parâmetros (HARTLEY, 1964; MATTHEWS, 1988; DOLAN et al., 2007; VUGRIN et al., 2007). Além do mais, ter o parâmetro de interesse no modelo confere maior interpretabilidade ao mesmo e permite extensões com relação à métodos de inferência e especificação de modelos.

Tabela 1 Estimativas, limites inferiores (LI) e limites superiores (LS) para o intervalos de confiança assintóticos (95%) para os parâmetros nas duas parametrizações do modelo potência para cada estágio fenológico do algodoeiro

Estágio	Parâmetro	Original			Reparametrizado		
		Est.	LI	LS	Est.	LI	LS
Vegetativo	θ_0	31,16	29,26	33,06	31,16	29,26	33,06
	θ_1	9,29	6,00	12,59	9,29	6,00	12,59
	θ_c	1,52	0,59	2,45			
	ϑ_q	0,87	0,73	1,02	0,87	0,73	1,02
Botão floral	θ_0	29,27	27,26	31,27	29,27	27,26	31,27
	θ_1	5,70	2,44	8,96	5,70	2,44	8,96
	θ_c	1,28	-0,16	2,71			
	ϑ_q	0,96	0,79	1,14	0,96	0,79	1,14
Florescimento	θ_0	30,14	20,44	39,85	30,14	20,44	39,85
	θ_1	14,54	5,45	23,63	14,54	5,45	23,63
	θ_c	-0,62	-2,18	0,95			
	ϑ_q	0,14	-0,42	0,70	0,14	-0,42	0,70
Maçã	θ_0	28,27	26,52	30,02	28,27	26,52	30,02
	θ_1	19,72	16,67	22,76	19,72	16,67	22,76
	θ_c	1,54	1,13	1,94			
	ϑ_q	0,74	0,64	0,85	0,74	0,64	0,85
Capulho	θ_0	34,79	30,38	39,20	34,79	30,38	39,20
	θ_1	8,79	4,43	13,14	8,79	4,43	13,14
	θ_c	-0,40	-1,68	0,87			
	ϑ_q	0,43	-0,27	1,13	0,43	-0,27	1,13

Valores para ϑ_q no modelo original foram obtidos pelo método delta.

Considerou-se $q = 5$ g.

Ao avaliar a extensão da não linearidade do modelo nas duas parametrizações pelas medidas de curvatura de Bates e Watts (1980), verificou-se que a curvatura devido à parametrização (CP) do modelo reparametrizado apresentou valores inferiores ao original para os estágios vegetativo e maçã. Nesses estágios os intervalos de confiança para θ_c não contiveram o valor zero, indicando uma relação não linear (Tabela 2). Nos estágios de florescimento e capulho, a CP para o modelo reparametrizado foi por volta de duas vezes maior que o modelo original. Nesses dois estágios, a estimativa de θ_c foi negativa e próxima a ponto de não diferir de zero. A curvatura intrínseca (CI) foi a mesma para ambas as parametrizações e os valores são aceitáveis, considerando 0,3 como ponto de corte (ROBINSON; HAMANN, 2011).

Tabela 2 Medidas de ajuste (R^2) e de curvatura devido ao efeito de parametrização (CP) e intrínseca (CI) para as parametrizações do modelo potência em cada estágio fenológico

Estágio	Devido à parametrização		Intrínseca	R^2
	Original	Reparametrizado		
Vegetativo	0,7045	0,5669	0,4258	0,6124
Botão floral	1,0356	3,8787	0,7334	0,3668
Florescimento	2,7293	7,6430	0,2255	0,4960
Maçã	0,3101	0,1989	0,1846	0,8929
Capulho	1,7184	2,2330	0,2425	0,5190

As medidas de curvatura para o modelo reparametrizado estão associadas aos valores de θ_c no modelo original, que tem correspondência com valores de ϑ_q no modelo reparametrizado. Sabe-se que se $\theta_c = 0$ então $\vartheta_q = e/\theta_1$, ou seja, a medida que o parâmetro θ_c se aproxima de zero, o valor de ϑ_q é uma função linear

de θ_1 . Com isso induz-se uma colinearidade entre as colunas da matriz gradiente,

$$\mathbf{F} = \frac{\partial f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \frac{\partial f(x_1, \boldsymbol{\theta})}{\partial \theta_1} & \dots & \frac{\partial f(x_1, \boldsymbol{\theta})}{\partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x_n, \boldsymbol{\theta})}{\partial \theta_1} & \dots & \frac{\partial f(x_n, \boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix}_{n \times p}, \quad (4.1)$$

que é usada para fazer inferências assintóticas, como obter os erros-padrões além de avaliar as medidas de curvatura do modelo. No caso de uma aproximada colinearidade em \mathbf{F} , a matriz de covariâncias de $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} = \hat{\sigma}^2(\mathbf{F}^\top \mathbf{F})^{-1}$, possui covariâncias altas que na escala de correlações são próximas de um. Verificou-se valores de correlação superiores a 0,95, em valor absoluto, entre os estimadores nos estágios botão floral, florescimento e capulhos (Tabela 3). Tais estágios foram justamente os casos em que θ_c não diferiu de zero e a CP foi superior àquela do modelo original. Esses resultados apontam para a existência de relação entre as medidas de curvatura e o grau de colinearidade da matriz \mathbf{F} , que aumenta à medida que θ_c tende à 0, indicando que um sub-modelo deve ser considerado.

Os gráficos de perfil de verossimilhança, tanto para o modelo original quanto para o modelo reparametrizado, apontaram limites não finitos para os intervalos de confiança (95%) baseados na função deviance para pelo menos um dos parâmetros (Figuras 3 e 4). Tais resultados sugerem que o valor da verossimilhança não é reduzido o suficiente ao ponto de alcançar o valor de corte na deviance para encontrar limites do intervalo de confiança. Nesse caso, os intervalos apresentam algum dos extremos no infinito ou na borda do espaço paramétrico. Esses resultados sugerem que um modelo mais simples que o de três parâmetros deveria ser pode ser considerado. Pelo estudo dos sub-modelos gerados a partir do modelo potência, sabe-se que valores de θ_c indo para algum dos extremos da reta correspondem à valores de ϑ_q indo para algum dos extremos do intervalo (0,1) e nessas

Tabela 3 Variâncias (diagonal) e correlações (abaixo da diagonal) assintóticas entre os estimadores dos parâmetros do modelo reparametrizado para cada estágio fenológico

Estágio	Parâmetro	θ_0	θ_1	ϑ_q
Vegetativo	θ_0	0,9408		
	θ_1	0,6377	2,8260	
	ϑ_q	-0,7111	-0,7166	0,0057
Botão floral	θ_0	1,0456		
	θ_1	0,6382	2,7655	
	ϑ_q	-0,7348	-0,9606	0,0080
Florescimento	θ_0	24,5121		
	θ_1	0,9322	21,5158	
	ϑ_q	-0,9550	-0,8558	0,0809
Maçã	θ_0	0,7997		
	θ_1	0,6379	2,4204	
	ϑ_q	-0,6543	-0,5586	0,0027
Capulho	θ_0	5,0680		
	θ_1	0,8780	4,9378	
	ϑ_q	-0,9552	-0,8436	0,1268

situações um sub-modelo é obtido.

O estágio de maçã, no entanto, foi uma exceção, pois apresentou limites finitos e formato simétrico para o perfil. Isso está ligado às baixas medidas de curvatura e ao fato de que é mais pronunciada a relação não linear entre produção e desfolha, de acordo com o modelo potência, no estágio de maçã se comparado aos demais.

Motivado pelas evidências levantadas, foram ajustados dois sub-modelos aos dados: o modelo (2.7), que é uma parametrização do modelo linear, e o modelo nulo com apenas intercepto (θ_0). Comparações com o modelo de três parâmetros foram feitas pelo teste F baseado na redução em soma de quadrados entre modelos aninhados que é equivalente ao teste de razão de verossimilhanças. Verificou-se que nos estágios de botão floral, florescimento e capulho, o modelo linear é tão adequado quanto o modelo potência (Tabela 4). O modelo potência deve ser mantido para os estágios vegetativo e maçã. Justamente nesses estágios, obteve-se baixas medidas de curvatura, o que permite concluir que a parametrização tem melhores propriedades inferenciais em situações em que a relação produção-desfolha é pronunciadamente não linear. Nos demais estágios, um modelo mais simples deve ser considerado.

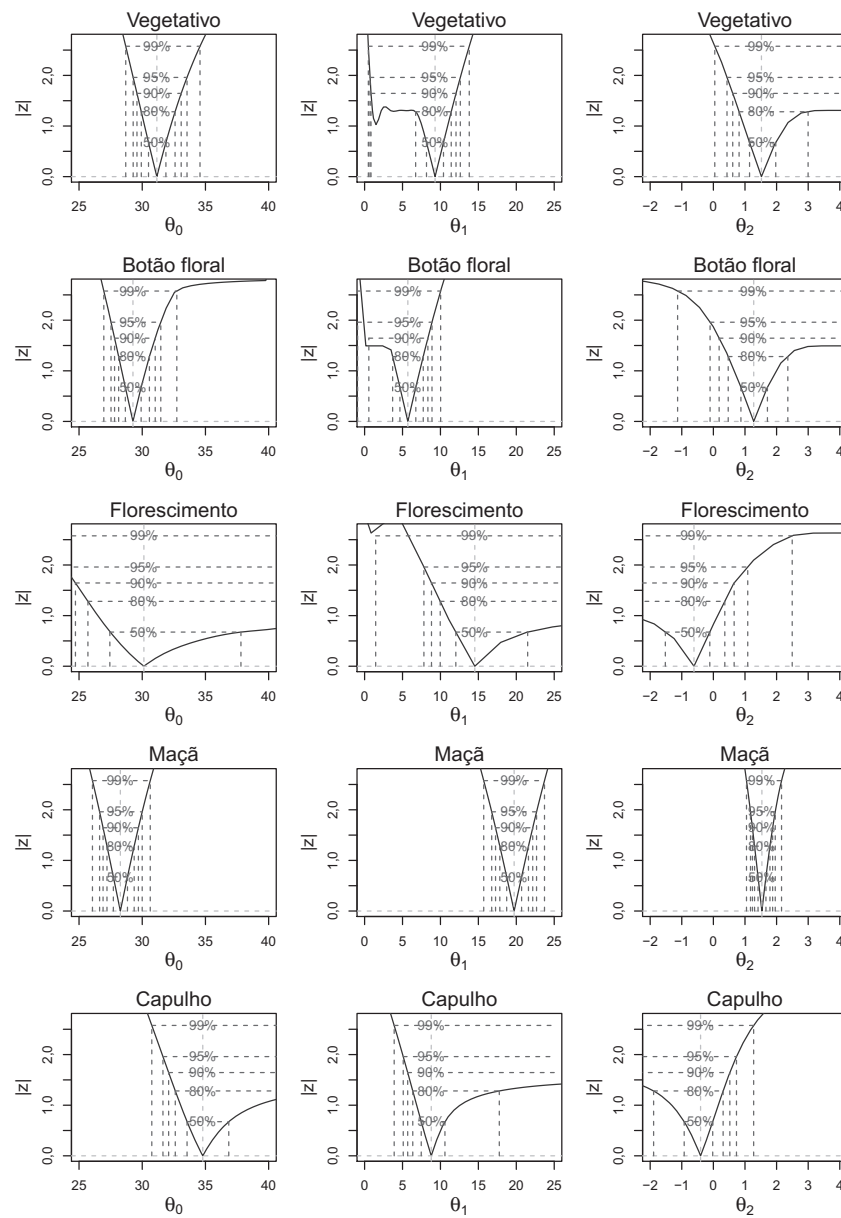


Figura 3 Gráficos dos perfis de verossimilhança para os parâmetros do modelo original (colunas) para cada estágio fenológico (linhas)

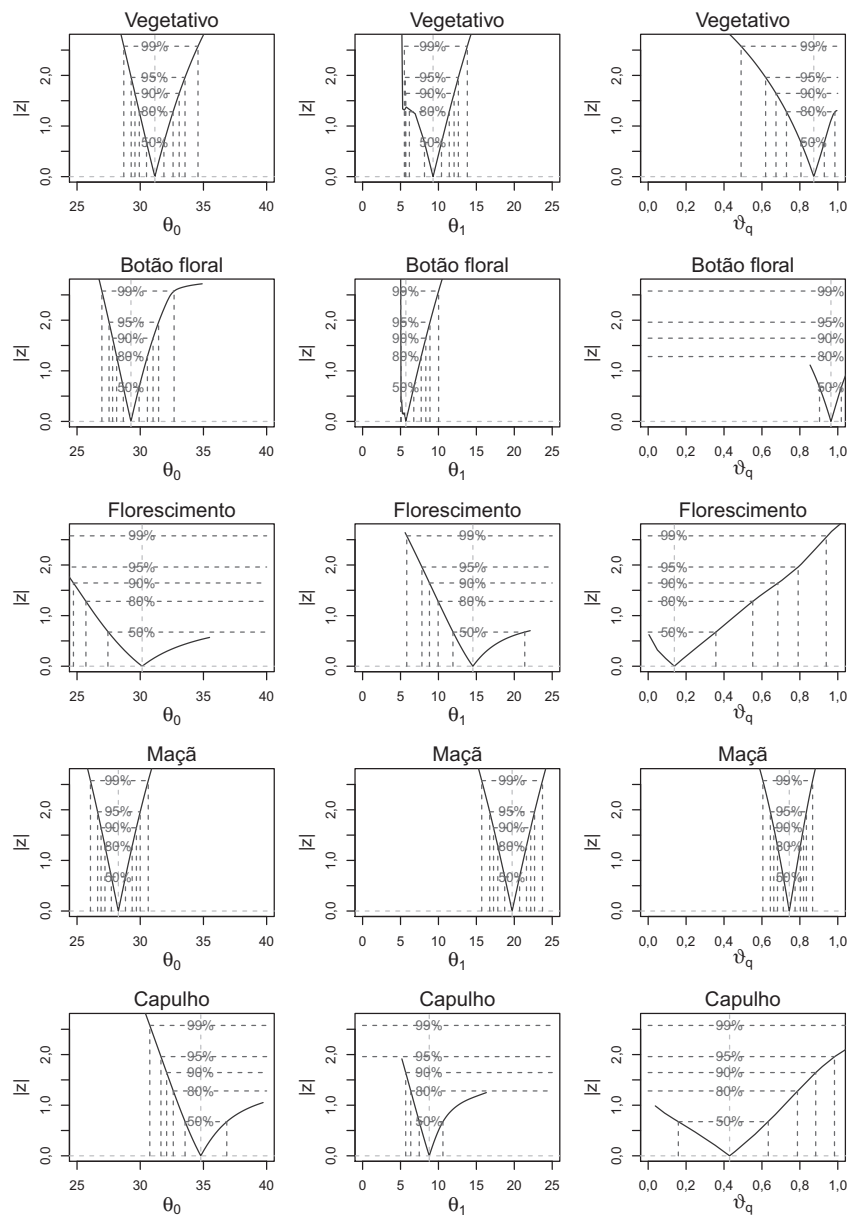


Figura 4 Gráficos dos perfis de verossimilhança para os parâmetros do modelo reparametrizado (colunas) para cada estágio fenológico (linhas)

Tabela 4 Resumo do teste F sobre modelos aninhados para cada estágio fenológico

Estágio	Modelo	GLR	SQR	DGL	DSQ	F	valor p
Vegetativo	Potência	22	238,3				
	Linear	23	316,3	1	-78,0	7,2	0,014
	Nulo	24	614,9	1	-298,6	21,7	0,000
Botão floral	Potência	22	230,6				
	Linear	23	265,1	1	-34,5	3,3	0,083
	Nulo	24	364,1	1	-99,1	8,6	0,007
Florescimento	Potência	22	520,5				
	Linear	23	534,9	1	-14,4	0,6	0,444
	Nulo	24	1032,9	1	-497,9	21,4	0,000
Maçã	Potência	22	204,1				
	Linear	23	700,7	1	-496,7	53,5	0,000
	Nulo	24	1905,9	1	-1205,1	39,6	0,000
Capulho	Potência	22	193,6				
	Linear	23	197,6	1	-4,0	0,5	0,509
	Nulo	24	402,5	1	-204,9	23,9	0,000

GLR: graus de liberdade dos resíduos; SQR: soma de quadrados dos resíduos;

DGL: diferença em graus de liberdade; DSQ: diferença em soma de quadrados.

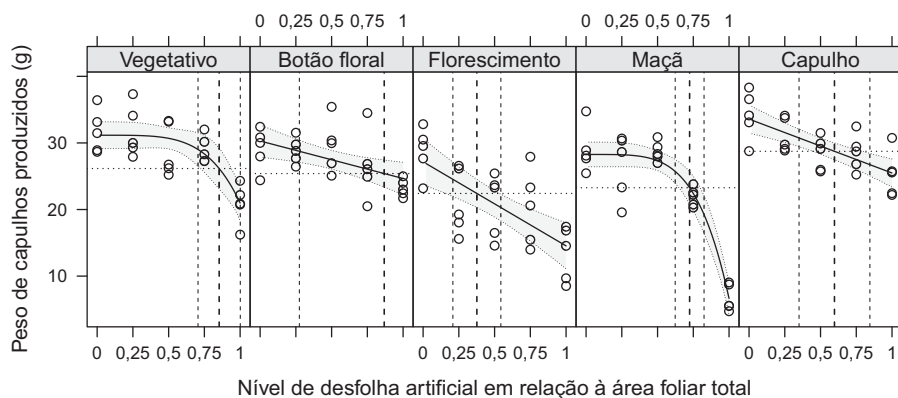


Figura 5 Valores observados e preditos com banda de confiança assintótica (95%) para o peso de capulhos produzido em função do nível de desfolha artificial em cada estágio fenológico do algodoeiro. Linhas verticais identificam a estimativa intervalar assintótica (95%) para o nível de dano econômico θ_q da desfolha na produção e linhas horizontais representam a produção 5 g abaixo da produção normal

Tabela 5 Estimativas, limites inferiores (LI) e limites superiores (LS) para o intervalo de confiança assintótico (95%), para os parâmetros sob o modelo final em cada estágio fenológico

Estágio	Parâmetro	Estimativa	LI	LS
Vegetativo	θ_0	31,157	29,256	33,058
	θ_1	9,291	5,997	12,586
	ϑ_q	0,873	0,726	1,021
Botão floral	θ_0	30,403	28,036	32,770
	ϑ_q	0,888	0,294	1,482
Florescimento	θ_0	27,416	24,054	30,778
	ϑ_q	0,396	0,228	0,564
Maçã	θ_0	28,269	26,516	30,022
	θ_1	19,715	16,666	22,764
	ϑ_q	0,744	0,643	0,845
Capulho	θ_0	33,743	31,700	35,786
	ϑ_q	0,617	0,370	0,865

Pelo teste F verifica-se que as medidas de curvatura elevadas, as altas correlações da matriz de covariâncias e principalmente os gráficos de perfil de verossimilhança com limites não finitos, sugerem que um modelo mais simples deveria ser considerado.

Valores preditos para os modelos selecionados, de acordo com o teste F, estão representados na figura 5, juntamente com a banda de confiança (95%) e com destaque para a estimativa intervalar assintótica (95%) para o nível de dano econômico ϑ_q . O algodoeiro é mais sensível à leves desfolhas no estágio de florescimento, com um padrão de decaimento linear no qual uma desfolha de 0,396 corresponde à dano econômico (Tabela 5). Nos demais estágios $\hat{\vartheta}_q$ foi superior a 0,50. O decaimento também é linear no botão floral e capulho, no entanto com taxa de variação inferior ao florescimento, com níveis de desfolha de dano econômico iguais a 0,888 e 0,617, respectivamente. No estágio vegetativo e de maçã a relação

foi côncava, sendo que a produção praticamente não diminuiu até 0,5 de desfolha e o dano econômico corresponde as desfolhas de 0,873 e 0,744, respectivamente.

5 CONCLUSÕES

O propósito da reparametrização foi representar o nível de dano econômico no modelo. As parametrizações foram comparadas com relação aos métodos disponíveis para fazer inferência sobre o nível de dano econômico. A inferência baseada em verossimilhança foi mais adequada no sentido de auxiliar a seleção de modelos. Além do mais, verificou-se que as medidas de curvatura e inspeção da matriz de covariâncias das estimativas também são úteis no processo de seleção de modelos. Nos estágios com pronunciada relação linear para produção-desfolha, o modelo reparametrizado apresentou melhores propriedades inferenciais. O algodoeiro responde de forma diferenciada à desfolha em cada estágio fenológico.

REFERÊNCIAS

- ALPER, J. S.; GELB, R. I. Standard errors and confidence intervals in nonlinear regression: comparison of Monte Carlo and parametric statistics. **The Journal of Physical Chemistry**, Washington, v. 94, n. 11, p. 4747–4751, May 1990.
- BATES, D.; WATTS, D. Relative curvature measures of nonlinearity. **Journal of the Royal Statistical Society. Series B (Methodological)**, London, v. 42, n. 1, p. 1–25, Jan. 1980.
- BATTAGLIA, M. et al. Modelling the impact of defoliation and leaf damage on forest plantation function and production. **Ecological Modelling**, Amsterdam, v. 222, n. 17, p. 3193–3202, Sept. 2011.
- BEALE, E. M. Confidence regions in non-linear estimation. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 22, n. 1, p. 41–76, Jan. 1960.
- BENDER, R. Calculating confidence intervals for summary measures of individual curves via nonlinear regression models. **International Journal of Bio-Medical Computing**, Limerick, v. 41, n. 1, p. 13–18, Mar. 1996.
- BERTONCELLO, T. F. et al. Desfolha artificial simulando danos de pragas no cultivo de arroz de sequeiro de casa de vegetação. **Revista Acadêmica: ciências agrárias e ambientais**, Curitiba, v. 9, n. 1, p. 33–39, Mar. 2011.
- CHENG, R.; EVANS, B.; ILES, T. Embedded models in non-linear regression. **Journal of the Royal Statistical Society. Series B (Methodological)**, London, v. 54, n. 3, p. 877–888, Aug. 1992.
- COUGHENOUR, M. et al. Production and nitrogen responses of the African dwarf shrub *Indigofera spinosa* to defoliation and water limitation. **Oecologia**, Berlin, v. 83, n. 1, p. 546–552, July 1990.
- COX, C. Fieller's theorem, the likelihood and the delta method. **Biometrics**, Washington, v. 46, n. 3, p. 709–718, Dec. 1990.
- DOLAN, K.; YANG, L.; TRAMPEL, C. Nonlinear regression technique to estimate kinetic parameters and confidence intervals in unsteady-state conduction-heated foods. **Journal of Food Engineering**, Essex, v. 80, n. 2, p. 581–593, May 2007.

- EDELSTEN, P. R.; CORRALL, A. J. Regression models to predict herbage production and digestibility in a non-regular sequence of cuts. **The Journal of Agricultural Science**, Cambridge, v. 92, n. 03, p. 575–585, June 1979.
- FAIRCLOTH, J. C. et al. The influence of defoliation timing on yields and quality of two cotton cultivars. **Crop Science**, Madison, v. 44, n. 1, p. 165–172, Jan. 2004.
- FAIRCLOTH, J. C. et al. Timing defoliation applications for maximum yields and optimum quality in cotton containing a fruiting gap. **Crop Science**, Madison, v. 44, n. 1, p. 158–164, Jan. 2004.
- GUTIERREZ, A. P. et al. An analysis of cotton production in California: a model for acala cotton and the effects of defoliators on its yields. **Environmental Entomology**, College Park, v. 4, n. 1, p. 125–136, Feb. 1975.
- HARE, D. J. Impact of defoliation by the colorado potato beetle on potato yields. **Journal of Economic Entomology**, v. 73, n. 3, p. 369–373, jun. 1980.
- HARTLEY, H. Exact confidence regions for the parameters in non-linear regression laws. **Biometrika**, London, v. 51, n. 3/4, p. 347–353, Dec. 1964.
- KAWAMOTO, H.; SAITO, T.; KIRITANI, K. A soybean growth model based on compensatory growth following defoliation. **Ecological Research**, Tsukuba, v. 1, n. 2, p. 195–206, Aug. 1986.
- KLUBERTANZ, T. H.; PEDIGO, L. P.; CARLSON, R. E. Soybean physiology, regrowth, and senescence in response to defoliation. **Agronomy Journal**, Madison, v. 88, n. 4, p. 577, July 1996.
- KULMAN, H. M. Effects of insect defoliation on growth and mortality of trees. **Annual Review of Entomology**, Stafsord, v. 16, n. 1, p. 289–324, Jan. 1971.
- LAUER, J. G.; ROTHB, G. W.; BERTRAMC, M. G. Impact of defoliation on corn forage yield. **Agronomy Journal**, Madison, v. 96, n. 5, p. 1459–1463, Sept. 2004.
- Lima Junior, I. et al. Desfolha artificial simulando danos de pragas na cultura do girassol (*Helianthus annuus* L., Asteraceae). **Revista Ceres**, Lavras, v. 57, n. 1, p. 23–27, 2010.
- LUDWIG, L.; SAEKI, T.; EVANS, L. Photosynthesis in artificial communities of cotton plants in relation to leaf area I: experiments with progressive defoliation of

mature plants. **Australian Journal of Biological Sciences**, Meuborne, v. 18, n. 6, p. 1103–1118, May 1965.

MATTHEWS, D. Likelihood-based confidence intervals for functions of many parameters. **Biometrika**, London, v. 75, n. 1, p. 139–144, Sept. 1988.

MOTAZEDIAN, I.; SHARROW, S. H. Defoliation frequency and intensity effects on pasture forage quality. **Journal of Range Management**, Denver, v. 43, n. 3, p. 198–201, May 1990.

PAINTER, E. L.; DETLING, J. K. Effects of defoliation on net photosynthesis and regrowth of western wheatgrass. **Journal of Range Management**, Denver, v. 34, n. 1, p. 68–71, Jan. 1981.

ROBINSON, A. P.; HAMANN, J. D. **Forest analytics with R**. 1. ed. New York: Springer, 2011. 354 p.

SCHAAFSMA, A.; ABLETT, G. Yield loss response of navy bean to partial or total defoliation. **Journal of Production Agriculture**, v. 7, n. 2, p. 202–206, June 1994.

SCHABENBERGER, O.; PIERCE, F. J. **Contemporary statistical models for the plant and soil Sciences**. 1. ed. Boca Raton: CRC, 2002. 738 p.

SILVA, A. M. et al. Impacto de diferentes níveis de desfolha artificial nos estádios fenológicos do algodoeiro. **Revista de Ciências Agrárias**, Belém, v. 35, n. 1, p. 163–172, June 2012.

VOLESKY, J. D. et al. Defoliation effects on herbage production and root growth of wet meadow forage species. **Rangeland Ecology & Management**, v. 64, n. 5, p. 506–513, Sept. 2011.

VUGRIN, K. W. et al. Confidence region estimation techniques for nonlinear regression in groundwater flow: Three case studies. **Water Resources Research**, Washington, v. 43, n. 3, p. n/a–n/a, Mar. 2007.

WACHENDORF, M. et al. Performance and environmental effects of forage production on sandy soils. II. Impact of defoliation system and nitrogen input on nitrate leaching losses. **Grass and Forage Science**, Oxford, v. 59, n. 1, p. 56–68, Mar. 2004.

ZEHNDER, G. W.; EVANYLO, G. K. Influence of Colorado potato beetle sample counts and plant defoliation on potato tuber production. **American Potato Journal**, Orono, v. 65, n. 12, p. 725–736, Dec. 1988.

ZEWDU, T. Effect of defoliation and intercropping with forage legumes on maize yield and forage production. **Tropical Science**, London, v. 43, n. 4, p. 204–207, Dec. 2003.

ANEXOS

ANEXO A Código R reproduzível correspondente ao ajuste do modelo potência reparametrizado para inferência sobre o nível de dano econômico da desfolha no algodoeiro. Disponível online em: <http://www.leg.ufpr.br/~walmes/TESE/anexoDESF.R>

```
##=====
## Uma parametrização de um modelo não linear para inferência
## sobre o nível de dano econômico da desfolha no algodoeiro
##
##                                     Walmes Marques Zeviani
##                                     walmes@ufpr.br
##
##=====

##-----
## Definições da sessão.

require(lattice)
require(car)
require(bbmle)
require(plyr)
require(reshape)

##-----
## Carrega os dados.

url <- "http://www.leg.ufpr.br/~walmes/data/ZevianiTeseDESF.txt"
da <- read.table(url, header=TRUE, sep="\t")
da$estágio <- factor(da$estágio,
                    levels=c("Vegetativo", "Botão floral",
                              "Florescimento", "Maça", "Capulho"))

str(da)

##-----
## Visualiza os dados.

xyplot(peso~desfolha|estágio, data=da)

##-----
## Define funções que retornam F e H para obter medidas de
## curvatura.

## modelo potência original
m.0 <- deriv3(~theta0-theta1*x^exp(thetac),
              c("theta0", "theta1", "thetac"),
              function(x, theta0, theta1, thetac){ NULL })
```

```

## modelo potência reparametrizado para dano econômico
## valor de q=5, mudar se for o caso dentro da fórmula
m.DE <- deriv3(~theta0-theta1*x^((log(5)-log(theta1))/log(varthetaq)),
               c("theta0","theta1","varthetaq"),
               function(x, theta0, theta1, varthetaq){ NULL })

##-----
## Ajusta as parametrizações separado por estágio fenológico.

da.e <- split(da, f=da$estágio)
str(da.e)

a0 <- lapply(da.e, nls,
             formula=peso~m.0(desfolha+0.02,
                               theta0, theta1, thetac),
             start=list(theta0=30, theta1=8, thetac=1))
lapply(a0, summary)

aDE <- lapply(da.e, nls,
             formula=peso~m.DE(desfolha+0.02,
                                theta0, theta1, varthetaq),
             start=list(theta0=30, theta1=8, varthetaq=0.7))
lapply(aDE, summary)

##-----
## Medidas de curvatura.

lapply(a0, MASS::rms.curv)
lapply(aDE, MASS::rms.curv)

##-----
## Matriz de correlação entre estimativas

lapply(a0, function(m) cov2cor(vcov(m)))
lapply(aDE, function(m) cov2cor(vcov(m)))

##-----
## Inferência sobre o dano econômico pelo método delta.

B <- coef(a0[[1]])
W <- vcov(a0[[1]])
g <- "(5/theta1)^exp(-thetac)"

deltaMethod(object=B, g=g, vcov=W)

dm <- function(nlsobj){
  B <- coef(nlsobj); W <- vcov(nlsobj)
  dm <- unlist(deltaMethod(object=B, g=g, vcov=W))
  dm <- c(dm, lwr=dm[1]-1.96*dm[2], upr=dm[1]+1.96*dm[2])
  names(dm) <- gsub("\\..*", "", names(dm))
  return(dm)
}

```

```

}

varthetaq <- lapply(a0, dm)
varthetaq <- ldply(varthetaq)
varthetaq

##-----
## Funções de verossimilhança.

ll0 <- function(theta0, theta1, thetac, desfolha, peso){
  x <- desfolha+0.02; y <- peso
  ex <- theta0-theta1*x^exp(thetac)
  sd <- sqrt(crossprod(y-ex)/(length(ex)))
  ll <- sum(dnorm(y, mean=ex, sd=sd, log=TRUE))
  -ll
}

llDE <- function(theta0, theta1, varthetaq, desfolha, peso){
  x <- desfolha+0.02; y <- peso
  ex <- theta0-theta1*x^((log(5)-log(theta1))/log(varthetaq))
  sd <- sqrt(crossprod(y-ex)/(length(ex)))
  ll <- sum(dnorm(y, mean=ex, sd=sd, log=TRUE))
  -ll
}

##-----
## Ajuste dos modelos.

a0 <- alDE <- da.e
for(i in names(a0)){
  a0[[i]] <- mle2(minuslogl=ll0,
                 start=as.list(coef(a0[[i]])),
                 data=c(as.list(subset(da, estágio==i))),
                 method="BFGS")
  alDE[[i]] <- mle2(minuslogl=llDE,
                   start=as.list(coef(alDE[[i]])),
                   data=c(as.list(subset(da, estágio==i))),
                   method="BFGS")
}

##-----
## Estimativas e medidas de ajuste.

cbind(t(sapply(a0, coef)), t(sapply(alDE, coef)))
cbind(sapply(a0, logLik), sapply(alDE, logLik))

##-----

par(mfrow=c(5,3), mar=c(2.2,2.2,1,1))
lapply(a0, function(m) plot(profile(m, maxsteps=20)))
layout(1)

```

```
par(mfrow=c(5,3), mar=c(2.2,2.2,1,1))
lapply(alDE, function(m) plot(profile(m, maxsteps=20)))
layout(1)
```

```
##-----
```

de algodão em ANEXO B Peso de capulhos produzido (g) ao final do ciclo a cada duas plantas de algodão em função do nível de desfolha artificial e estágio fenológico.

Estágio	Desfolha	Peso de capulhos produzido (g)				
Vegetativo	0,00	33,160	28,675	31,485	28,925	36,430
Vegetativo	0,25	34,105	30,010	37,338	27,930	29,320
Vegetativo	0,50	25,228	26,205	33,327	26,750	33,200
Vegetativo	0,75	30,180	27,306	27,277	28,305	32,030
Vegetativo	1,00	24,323	20,875	20,755	16,235	22,200
Botão floral	0,00	24,400	30,015	27,980	32,437	30,845
Botão floral	0,25	28,815	31,547	26,453	29,775	27,700
Botão floral	0,50	30,389	26,957	29,975	35,430	25,093
Botão floral	0,75	34,501	26,860	20,500	26,071	24,929
Botão floral	1,00	23,015	25,007	21,706	22,449	24,024
Florescimento	0,00	27,681	30,510	29,555	32,825	23,200
Florescimento	0,25	26,580	15,600	19,260	26,174	18,060
Florescimento	0,50	23,666	16,491	23,310	14,590	25,425
Florescimento	0,75	27,928	15,475	23,314	20,620	13,990
Florescimento	1,00	16,851	8,522	9,686	14,556	17,435
Maçã	0,00	28,103	28,900	25,460	27,690	34,750
Maçã	0,25	30,678	30,360	23,330	19,581	28,655
Maçã	0,50	27,991	26,994	30,878	28,358	29,390
Maçã	0,75	20,312	22,560	23,810	20,754	22,230
Maçã	1,00	8,750	4,735	5,560	9,055	5,535
Capulho	0,00	34,140	28,777	36,580	38,320	33,100
Capulho	0,25	29,128	34,105	28,870	29,760	33,763
Capulho	0,50	29,122	25,970	25,735	29,930	31,510
Capulho	0,75	29,460	25,255	32,480	26,820	28,740
Capulho	1,00	22,460	25,750	30,760	22,200	25,575

CAPÍTULO 3 PARAMETRIZAÇÃO DO MODELO VAN GENUCHTEN PARA INFERÊNCIA SOBRE OS PARÂMETROS S E I

RESUMO

A água é indispensável para produção das culturas pois está envolvida no transporte de nutrientes, reações químicas, processos físicos e manutenção da vida do solo. O conhecimento sobre a curva de retenção de água (CRA) do solo é fundamental para estabelecer estratégias de manejo. A qualidade física do solo é dependente da CRA e os parâmetros I , tensão do ponto de inflexão da CRA, e S , taxa de variação no ponto de inflexão, considerados como indicadores da qualidade física, são parâmetros relacionados a medidas descritivas da distribuição do tamanho de poros do solo. Com este trabalho, objetiva-se verificar o efeito da posição de amostragem e profundidade do solo sobre os parâmetros I e S da CRA. Para isso 1) considerou-se ANOVA simples e 2) ANOVA ponderada pela variância das estimativas desses parâmetros em cada unidade experimental em comparação com 3) o uso de modelos não lineares de efeito misto em uma parametrização desenvolvida para I e S . Nenhum dos métodos alternativos de análise foi superior ao modelo não linear de efeitos mistos na parametrização desenvolvida, que apresentou intervalos mais estreitos para estimativas dos parâmetros e apontou efeito de posição e profundidade de coleta nos parâmetros I e S .

Palavras-chave: Reparametrização. Função de parâmetros. Método delta. Curva de retenção de água.

ABSTRACT

Water is essential for crop production because it is involved in the transport of nutrients, chemical reactions and physical processes maintaining the soil life. The knowledge about the water retention curve (WRC) of the soil is important to establish management strategies. The physical quality of the soil is dependent on the WRC and the parameters I , the inflection point of the WRC, and S , the rate of change at the inflection point, considered as indicators of quality physical parameters are related to descriptive measures the pore size distribution of the soil. In this work, we aim to investigate the effect of sampling position and soil depth on the parameters I and S of the WRC. For that 1) it was considered ANOVA and 2) weighted ANOVA based on the variance of these estimates in each experimental unit compared to 3) using nonlinear mixed-effects in a parameterization developed for I and S . None of the alternative methods of analysis was superior to the nonlinear mixed effects model in the parameterization developed, which showed narrower intervals for parameter estimates and pointed sampling position and depth effect on parameters I and S .

Key-words: Reparametrization. Function of parameters. Delta method. Water retention curve.

1 INTRODUÇÃO

Água é indiscutivelmente o fator isolado mais importante, com possível exceção para luz solar e o ar que não têm disponibilidade limitada, para o desenvolvimento das plantas (CHESWORTH, 2007). A água presente no solo desempenha inúmeras funções de ordem química, física e biológica como, por exemplo, dissolver e transportar nutrientes para as plantas além de também ser considerada um nutriente (BRADY; WEIL, 2009). O excesso de água no solo pode diminuir ou impedir o desenvolvimento das plantas devido a uma reduzida aeração, enquanto deficiências podem causar estresse hídrico e, se severa o suficiente, o murchamento e morte da planta. Para uma produção eficiente, um fornecimento constante de água é necessário durante o ciclo da planta para atender as necessidades hídricas e promover a dissolução, transporte e absorção de nutrientes. O conteúdo de água também influencia a qualidade de operações de manejo do solo e colheita.

A capacidade do solo em reter água é influenciada principalmente por sua textura, estrutura, conteúdo de matéria orgânica. Solos arenosos geralmente retêm pouca quantidade de água (menos de 10% em peso) na capacidade de campo. Por outro lado, solos argilosos (acima de 40% de argila em peso) retêm uma quantidade de água superior a 40%, em peso. A estrutura do solo, considerada como sendo a combinação e arranjo de partículas primárias (grânulos) em partículas secundárias de solo (agregados), influencia a retenção de água por meio da formação de poros. O grau de desenvolvimento estrutural do solo influencia o espaço ocupado por poros, que por sua vez influencia a quantidade de água que o solo pode reter.

Os poros são condição essencial de um solo. Um solo sem poros é considerado rocha. Os poros permitem que raízes e diversas formas de vida penetrem e se estabeleçam no solo. A infiltração e armazenamento de água, troca de gases e

transporte de substâncias são processos facilitados pela rede de poros que no total representa algo ao redor de 50% do volume do solo. Essa porosidade é distribuída ao longo de uma grande variação em termos de tamanho, frequência e forma que vão desde os microscópicos existentes entre lâminas dos minerais de argila até os grandes e contínuos poros deixados por minhocas e raízes decompostas.

A água é fixada no solo predominantemente por causa da sua atração à superfícies e as forças capilares associadas à tensão superficial, γ (N m^{-1}). Essas forças resultam em um fenômeno chamado de ascensão capilar que é a altura que a água apresenta dentro de um tubo capilar em relação ao nível do líquido e que depende do raio deste tubo. A altura (h , m) de ascensão da água em um tubo capilar uniforme, cilíndrico de raio r (m) é

$$h = \frac{2\gamma \cos(\alpha)}{\rho g r}, \quad (1.1)$$

em que ρ é a densidade da água (kg m^{-3}), g a aceleração da gravidade (m s^{-2}) e α é o ângulo de contato entre o líquido e a parede do tubo (Figura 1). Para estudos com poros do solo α é comumente considerado nulo. Quanto menor o raio maior é a ascensão capilar (Figura 1). A pressão ou potencial matricial (ψ , N m^{-2}) necessária para tornar essa água capilar em água livre, em termos de energia (N m) por unidade de volume (m^3) é

$$\psi = -\frac{2\gamma}{r}, \quad (1.2)$$

em que ψ está em unidade de pressão (N m^{-2}) (DEXTER, 2004). Assim, quanto maiores os poros no solo mais fácil é remover água deles. Os valores de ψ são negativos pois representam pressão de sucção mas por comodidade de explanação, trataremos ψ em valor absoluto.

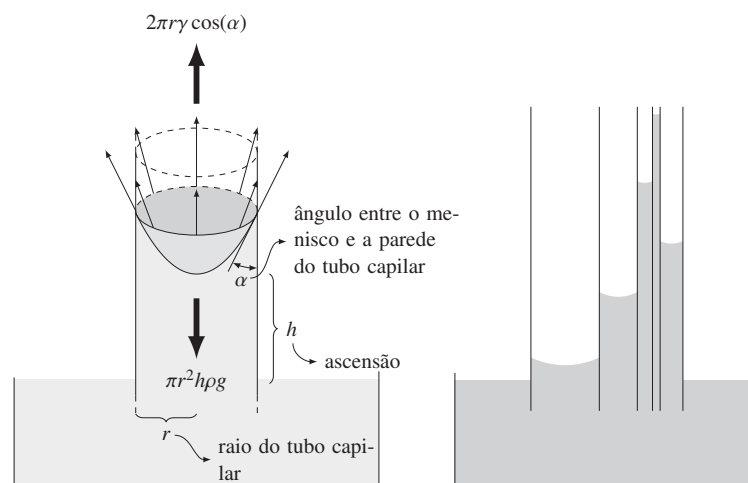


Figura 1 Equilíbrio entre as forças atuando sobre um líquido dentro de um tubo capilar (esq.) e relação entre raio do tubo capilar e altura de coluna líquido (dir.)

O conteúdo de água do solo (U) pode ser representado em termos de massa relativa ou base gravimétrica (kg de água por kg de solo) ou em termos de volume ou base volumétrica (m^3 de água por m^{-3} de solo). A diferença entre os valores nas duas escalas é apenas uma constante multiplicativa, a densidade do solo (kg m^{-3}) (MAIA, 2011). No solo, à medida que o ψ tende à zero, o conteúdo de água aumenta até que o solo se torna saturado (U_s) uma vez que todo espaço poroso fica ocupado por água. A água é removida de poros cada vez menores à medida que se aumenta ψ . Então existe uma relação entre o conteúdo de água e a tensão matricial.

A relação entre o conteúdo de água no solo e a tensão matricial é chamada de curva característica do solo ou curva de retenção de água do solo (CRA) (DEXTER; BIRD, 2001) e será representada por $U(\psi)$. Sua determinação é fundamental para conhecimento do comportamento hídrico do solo e decisões com

relação às estratégias de manejo (CARDUCCI et al., 2011) como, a avaliação da qualidade física do solo, balanço hídrico, operações de máquinas, zoneamento agrícola e previsão de produtividade.

Há vários modelos disponíveis para representar a CRA, no entanto, o modelo proposto por Genuchten (1980) tem sido adotado universalmente (OMUTO et al., 2006) pois é flexível ao ponto de representar bem e para uma ampla variedade de solos a relação entre conteúdo de água e tensão matricial. O modelo relaciona o conteúdo de água com a tensão matricial por

$$U(\psi) = U_r + \frac{U_s - U_r}{(1 + (A\psi)^n)^m}, \quad (1.3)$$

em que U_s e U_r são as assíntotas superior e inferior da CRA (m m^{-3} ou g g^{-1}), A (kPa^{-1}), n e m são adimensionais e considerados parâmetros empíricos do modelo. As modificações na CRA como função dos valores dos parâmetros está representada na figura 2. Apesar de ter cinco parâmetros, frequentemente adota-se o modelo sob a restrição de Mualem (1976) na qual $m = 1 - 1/n$ e adota-se a seguinte transformação de variável $A = \exp\{a\}$ por apresentar vantagens ao definir rotinas de otimização (OMUTO, 2007).

Dexter (2004) define o parâmetro S como sendo a inclinação tangente (S de *slope*) ao ponto de inflexão da CRA (U , kg kg^{-1}) na escala logarítmica da tensão. Em Dexter (2004), Dexter (2004b), Dexter (2004c) são apresentadas evidências do valor de S como indicador da qualidade física do solo e sugeridos intervalos de classificação da qualidade com relação à S . Nestes trabalhos, o modelo Genuchten (1980) foi considerado e expressões para cálculo de S foram obtidas como função dos parâmetros do modelo.

Após os trabalhos de Dexter (2004), Dexter (2004b), Dexter (2004c), extensa pesquisa foi realizada visando explorar e quantificar relações do parâmetro

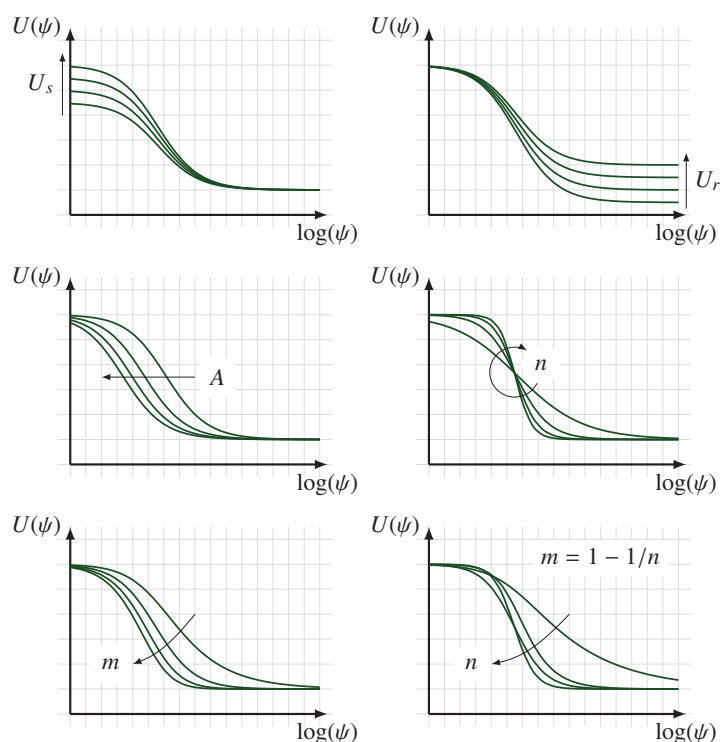


Figura 2 Curvas hipotéticas do modelo van Genuchten ($U, \log(\psi)$) com destaque para as modificações em relação aos valores de cada parâmetro. A seta indica a direção para incrementos regulares no valor do parâmetro

S com características dos solos. Tais pesquisas vão desde estudos de regressão e correlação (TORMENA et al., 2008; ANDRADE; STONE, 2009) à investigações a partir de experimentos à nível de campo onde as CRA são obtidas de acordo com uma estrutura de fatores experimentais como a classe de solo, o manejo e a profundidade do solo (MACHADO et al., 2008).

Nas pesquisas, as inferências à respeito de S são conduzidas em dois estágios: 1) ajuste da CRA para cada unidade amostral e determinação do valor de S e 2) correlação, regressão ou ANOVA considerando a estimativa de S como variável dependente e covariáveis ou fatores experimentais como variáveis explicativas.

Esse tipo de procedimento é universalmente aplicado e aceito dentro das pesquisas em física do solo. No entanto, para amostras pequenas como aquelas dos experimentos de campo ($n < 30$), tal procedimento tem demonstrado pouca capacidade para indicar relação de S com os fatores sob investigação.

A determinação da CRA envolve um processo laboratorial que consome tempo pois requer o secamento total seguido de saturação completa das amostras de solo que a partir de então, têm o conteúdo de água determinado para uma sequência de valores de tensão. Diante da velocidade com que certas decisões devem ser tomadas, frequentemente são usadas equações de predição para os parâmetros da CRA, chamadas funções de pedo-transferência, baseadas em medidas mais disponíveis como densidade e porosidade. Por outro lado, quando a determinação da CRA é feita, seus parâmetros são obtidos por ajuste de algum modelo teórico aos dados. No entanto, devido à fontes sistemáticas e aleatórias de variação, os modelos usadas não predizem exatamente os valores observados. Omutó et al. (2006) relata que diversas abordagens foram consideradas no sentido de melhorar a precisão do ajuste que vão desde a seleção de modelos, uso de melhores algoritmos de otimização, adoção de modelos semi-paramétricos e agrupamento de dados de acordo com tipo de solo e manejo, mas ressalva que embora as abordagens isoladamente forneçam o aprimoramento desejado, muitos pesquisadores não consideram as abordagens estatísticas que combinam seleção de modelos, algoritmos de otimização e uso de covariáveis no processo de estimação dos parâmetros.

O aumento considerável da capacidade computacional possibilita aplicação de métodos estatísticos que permitem aumento de acurácia para as estimativas dos parâmetros bem como incorporar efeito de covariáveis e informação de níveis de amostragem durante a análise. Os modelos não lineares de efeitos mistos

(NLMEM, *non linear mixed effects models*), que englobam os dados de medidas repetidas, permitem que variações nos dados possam ser explicadas por um modelo não linear além de acomodar os efeitos de grupos (tratamentos) e de informações ao nível de unidade experimental (covariáveis). Aplicativos para análise de NLMEM estão cada vez mais disponíveis e habilitados à propósitos gerais. As vantagens dos NLMEM incluem a capacidade de tratar dados desbalanceados e incompletos, flexibilidade para acomodar diversas fontes de variação e padrões de correlação e seu alto poder para comparar tratamentos (GREENLAND, 2000; QUENÉ; BERGH, 2004; ZHAO et al., 2005).

Os NLMEM tiveram grande difusão por meio de aplicações em estudos de farmacocinética onde são frequentes o registro de medidas de concentração-tempo de uma substância no sangue em que esse processo depende das características próprias do indivíduo avaliado (SHEINER; LUDDEN, 1992). Nas ciências agrárias um número crescente de aplicações de NLMEM pode ser encontrado (NIKO-LAIDIS et al., 1998; SCHABENBERGER; PIERCE, 2002) e nas ciências florestais aplicação predominante dos NLMEM é para aprimorar a predição de volume de madeira por incorporação de covariáveis no nível de indivíduo (NOTHDURFT et al., 2006), acomodação de efeito de indivíduos e de blocos (CALAMA; MONTERO, 2004; JORDAN et al., 2005; CALEGARIO et al., 2005; TEMESGEN et al., 2008), avaliação do efeito de tratamentos (ZHAO et al., 2005; WANG et al., 2007) e uso de estruturas de dependência serial (WANG et al., 2007; MENG; HUANG, 2010). Por outro lado, aplicações de NLMEM relacionados à CRA são escassos (OMUTO et al., 2006).

Para usar de todos os benefícios que os NLMEM apresentam ao se fazer inferências sobre determinado parâmetro é necessário que este faça parte do modelo não linear que representa a CRA. Uma vez que S é função dos parâmetros do

modelo Genuchten (1980), este modelo pode ser reparametrizado para estimação direta de S . Por meio de reparametrização, Rekaya et al. (2001) incorporaram o parâmetro persistência de lactação (P) ao modelo de Wood e com isso foram capazes de usar de informações de *pedigre* ao fazer estimação de P além de acomodar a variação entre animais e devido a ordem de lactação ao passo que Zeviani et al. (2012) estenderam e aplicaram reparametrização ao modelo exponencial assintótico para ter interpretação idêntica ao modelo Michaelis-Menten e assim selecionar o modelo com melhores propriedades do ponto de vista de inferência para representar a liberação de potássio no solo.

Neste trabalho considerou-se reparametrizações no modelo (GENUCHTEN, 1980) para representar a CRA como função de parâmetros diretamente ligados à qualidade física do solo. Ênfase foi dada ao processo de reparametrização. As medidas de não linearidade foram avaliadas para acessar a qualidade dessas reparametrizações. Modelos não lineares de efeitos mistos foram então considerados para acomodar o efeito aleatório das unidades amostrais e testar o efeito de tratamentos por meio da especificação de modelos. As inferências obtidas com os NLMEM foram comparadas com aquelas obtidas sob o procedimento de dois estágios. Na seção 2 é feita a conexão entre CRA e distribuição do tamanho de poros, na 3 os passos de reparametrização, na 4 o conjunto de dados é descrito bem como os métodos para análise, os resultados são apresentados na seção 5 com conclusões em 6.

2 RETENÇÃO DE ÁGUA E TAMANHO DE POROS

Dada a relação que existe entre tamanho de poros e tensão matricial (1.2) é possível chegar à distribuição de frequência de tamanho de poros quando o modelo Genuchten (1980) é usado para representar a CRA (REYNOLDS et al., 2002). Essa relação não pode ser considerada isoladamente pois fatores que alteram as propriedades da água de um solo para o outro a modificam, como a concentração de íons e a temperatura. Por outro lado, fixadas as demais variáveis, essa relação permite explorar a conexão existente entre a retenção de água e distribuição de tamanho de poros. Além do mais, a partir dessa relação, é possível verificar que o parâmetro S proposto por Dexter (2004) é o parâmetro de escala que representa o grau de dispersão de valores ao redor do tamanho modal de poros.

Para facilitar as operações algébricas, o modelo em (1.3) será escrito com a parametrização $\Delta = U_s - U_r$ e a mudança de variável $x = \log(\psi)$ usada por Dexter (2004), obtendo-se

$$U(x) = U_r + \frac{\Delta}{(1 + \exp\{a + x\}^n)^m}. \quad (2.1)$$

A derivada de $U(x)$ em relação à x , DRCA, representa a taxa instantânea de remoção de água que é diretamente proporcional à frequência de tamanho de poros (r) pois quanto mais água for removida entre dois limites de tensão maior é a proporção dos poros de tamanho correspondente, e é dada por

$$U'(x) = \frac{dU(x)}{dx} = -\Delta nm \exp\{n(a + x)\} (1 + \exp\{n(a + x)\})^{-m-1}. \quad (2.2)$$

Sabe-se por (2.2) que $\int_{-\infty}^{\infty} U'(x) = -\Delta$ e por (1.2) que $x = \log(\psi) = \log(2\gamma) - \log(r)$ e assim a função densidade de probabilidade da distribuição de tamanho de poros

(r) é dada por

$$f(r) = \frac{nm(2\gamma \exp\{a\}/r)^n}{r(1 + (2\gamma \exp\{a\}/r)^n)^{m+1}}, \quad (2.3)$$

é obtida ao padronizar $U'(x)$ dividindo por $-\Delta$. A função $f(r)$ satisfaz as condições de uma função de densidade de probabilidade e seus parâmetros são $-\infty < a < \infty$, $n > 0$ e $m > 0$.

Para se determinar o valor de S precisamos do valor da tensão no ponto de inflexão, I . Esse valor é a solução da equação obtida ao se igualar a segunda derivada de $U(x)$ à zero. A expressão da segunda derivada é

$$U''(x) = -\Delta mn^2 \exp\{n(a+x)\}(1 + \exp\{n(a+x)\})^{-m-1} \left(1 + \frac{(-m-1)\exp\{n(a+x)\}}{1 + \exp\{n(a+x)\}}\right) \quad (2.4)$$

e o ponto de inflexão é

$$I = -a - \log(m)/n. \quad (2.5)$$

Substituindo-se (2.5) em (2.2) tem-se

$$S = U'(I) = -\Delta n \left(1 + \frac{1}{m}\right)^{-m-1}. \quad (2.6)$$

Considere que X e Y representam as unidades de medida de x e $U(x)$, respectivamente. Sendo assim, pelo estudo de dimensionalidade, verifica-se que as unidade de medida de I , n e Δ são X , X^{-1} e Y . Portanto S tem unidade de medida $Y X^{-1}$, decorrente do termo Δn em (2.6), pois de fato S representa uma taxa. A Figura 3 ilustra a CRA com destaque para o significado gráfico dos parâmetros I e S também em termos da DRCA e da função de densidade do tamanho de poros. O parâmetro I representa a classe modal do tamanho de poros e S é um parâmetro de escala que representa a dispersão dos valores ao redor do valor modal. Resultados

semelhantes foram obtidos por Reynolds et al. (2009).

Reynolds et al. (2009) declara que o sucesso de S como índice que qualidade física e sua concordância com demais índices vem da relação que todos eles têm com a distribuição de tamanho de poros. Partindo desse ponto de vista, pode ser que I , tal como S , sirva como índice de qualidade física uma vez que representa uma medida de posição, a moda, da distribuição do tamanho de poros.

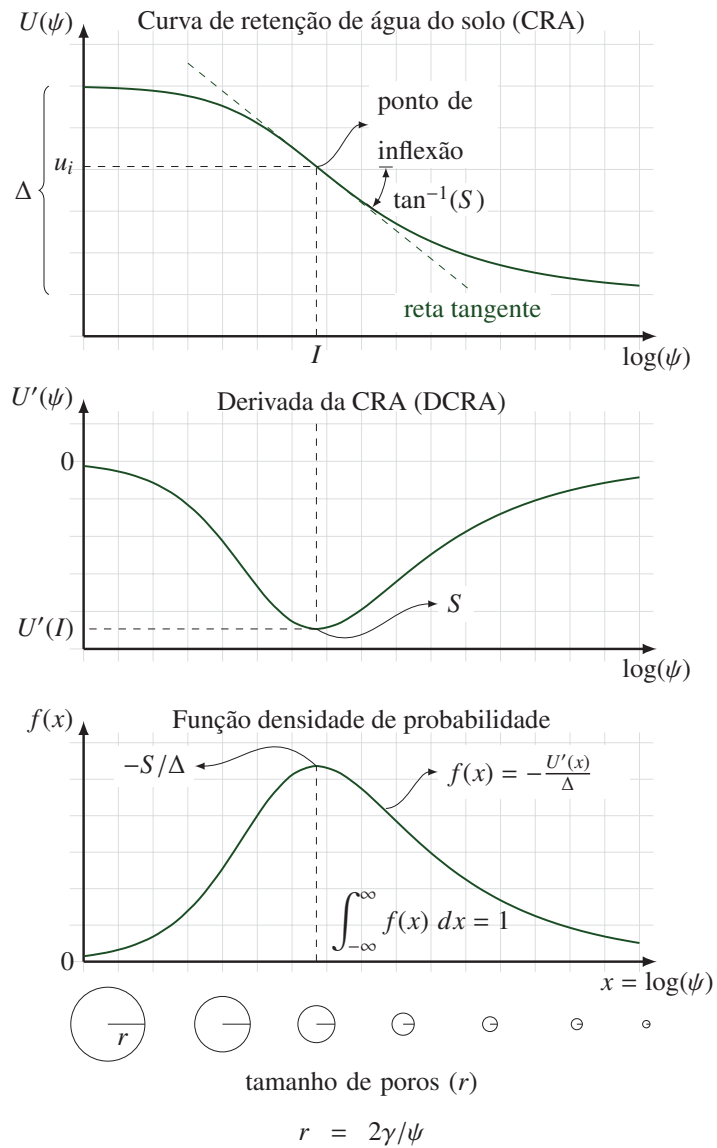


Figura 3 Curva de retenção de água do solo (topo), derivada da curva de retenção de água do solo (centro) e função densidade relacionada ao tamanho de poros do solo por meio da tensão matricial (base)

3 REPARAMETRIZAÇÃO

O modelo Genuchten (1980) será reparametrizado em termos de I e S . A definição de I e S está em 2.5 e 2.6. Pode-se isolar a como função de I por simplicidade e por terem a mesma unidade de medida, assim

$$a = -I - \log(m)/n. \quad (3.1)$$

Em contrapartida, o parâmetro S não possui uma alternativa tão óbvia pois o isolamento de Δ , U_r , U_s ou n são igualmente fáceis de obter

$$\Delta = -\frac{S(1 + 1/m)^{m+1}}{n}, \quad (3.2)$$

$$U_s = U_r - \frac{S(1 + 1/m)^{m+1}}{n}, \quad (3.3)$$

$$U_r = U_s + \frac{S(1 + 1/m)^{m+1}}{n}, \quad (3.4)$$

$$n = \frac{S(1 + 1/m)^{m+1}}{U_s - U_r}. \quad (3.5)$$

Dentre essas, a menos conveniente será n quando se assumir a restrição de Mualem (1976), $m = 1 - 1/n$, pois a expressão resultante não terá solução analítica, muito embora seja possível considerar uma solução numérica. Como é frequente a adoção da restrição de Mualem (1976) ênfase será dada nas opções restantes. Por fim, na etapa de substituição dessas expressões, após as simplificações convenientes, o modelo reparametrizado têm as seguintes expressões

$$U(x) = U_r - \frac{S(1 + 1/m)^{m+1}}{n(1 + \exp\{n(x - I)\}/m)^m}, \quad \text{em } \Delta \text{ ou } U_s,$$

$$U(x) = U_s - S(1 + 1/m)^{m+1}/n - \frac{2U_s + S(1 + 1/m)^{m+1}/n}{(1 + \exp\{n(x - I)\}/m)^m}, \quad \text{em } U_r.$$

Nos modelos resultantes tem-se que

$U(x)$ é a umidade (Y) para valores de tensão matricial $x = \log(\psi)$ (X);

U_r é a umidade residual (Y) pois representa o valor limite da umidade quando a tensão matricial ψ vai para o infinito, ou seja, $\lim_{\psi \rightarrow \infty} U(\psi) = U_r$;

U_s é a umidade de saturação (Y) pois representa o valor limite da umidade quando a tensão matricial ψ vai para zero, ou seja, $\lim_{\psi \rightarrow 0} U(\psi) = U_s$;

S é a inclinação da reta tangente ($Y X^{-1}$) no ponto de inflexão, ou seja, $S = U'(I)$. Por definição S representa a maior taxa de retenção de água do solo em relação à escala de tensão matricial $x = \log(\psi)$, $S < 0$;

I é a tensão no ponto de inflexão (X), $-\infty < I < \infty$. Associado a essa tensão está a classe de tamanho de poro de maior frequência, portanto I representa a tensão correspondente ao tamanho modal de poros;

n é chamado de parâmetro empírico do modelo pois não tem interpretação direta. Graficamente, alterações em n alteram a taxa de retenção de água, $n > 0$ (X^{-1});

m é um parâmetro empírico do modelo, $m > 0$. Sob a restrição de Mualem (1976), m é função de n , $m = 1 - 1/n$ e dessa forma $n > 1$;

Entre os modelos obtidos, aquele resultante da reparametrização em U_r apresenta uma expressão longa e como qualquer um deles é igualmente útil, pois incluem S e I , preferência será dada a

$$U(x) = U_r - \frac{S}{n} \frac{(1 + 1/m)^{m+1}}{(1 + \exp\{n(x - I)\}/m)^m}. \quad (3.6)$$

A Figura 4 apresenta as modificações na forma da CRA devidas aos parâmetro do modelo van Genuchten reparametrizado. A figura destaca que U_r controla a altura da curva e funciona como um intercepto; S controla a inclinação no ponto de inflexão e para isso deve ocorrer um aumento do valor da assíntota

superior que no modelo original é U_s ; I causa translação horizontal do ponto de inflexão; n modifica a curva em termos de forma e assíntota superior. Além do mais, verificam-se modificações semelhantes na CRA com relação às alterações de n e m . Então entende-se que a presença do parâmetro m confere pouca flexibilidade adicional e que a restrição de Mualem (1976) é apropriada. Com isso, o modelo tem a seguinte expressão

$$U(x) = U_r - \frac{S}{n} \frac{\left(\frac{2n-1}{n-1}\right)^{2-1/n}}{\left(1 + \left(\frac{n}{n-1}\right) \exp\{n(x - I)\}\right)^{1-1/n}}. \quad (3.7)$$

O fato de U_r e S serem parâmetros que ocorrem linearmente no modelo oferece vantagens computacionais relacionadas à estimação que pode ser numérica para os parâmetros n e I e analítica em U_r e S (VENABLES; RIPLEY, 2002). Isso reduz o tempo computacional pois a procura se dá em duas, em vez de quatro dimensões. Do ponto de vista de especificação de modelos, é mais conveniente computacionalmente associar efeitos aleatórios a parâmetros que ocorrem de forma linear (PINHEIRO; BATES, 2000).

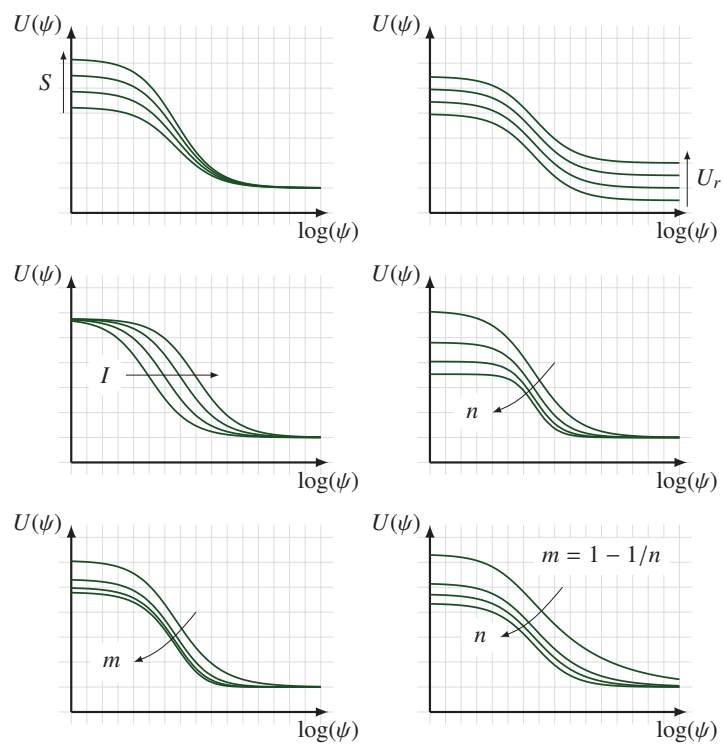


Figura 4 Curvas hipotéticas do modelo van Genuchten reparametrizado com destaque para as modificações em relação aos valores de cada parâmetro. As setas indicam as modificações para aumento equidistante no valor dos parâmetros

4 MATERIAL E MÉTODOS

Os dados considerados são medidas de conteúdo de água do solo ($\text{m}^3 \text{m}^{-3}$) em função do potencial matricial (kPa) coletados em campo sob cultura do café no qual se utilizam métodos conservacionistas e de manejo intensivo do solo (SERAFIM et al., 2011). Amostras de solo indeformadas foram coletadas de um solo classificado como Latossolo Vermelho distroférico (LVd) em uma lavoura com 3,5 anos de implantação.

As amostras estão vinculadas à uma estrutura experimental que considerou 2 níveis de posição de amostragem (linha e entre linha) da cultura e profundidade da camada do solo (0-20, 20-40, 40-80, 80-120, 120-160 cm), com três repetições, perfazendo 30 unidades experimentais. As tensões aplicadas às amostras foram 0, 1, 2, 4, 6, 10, 33, 100, 500 e 1500 kPa, resultando em 300 pares de valores de conteúdo de água (U) e tensão matricial (ψ). A figura 5 apresenta o gráfico de dispersão das observações.

O objetivo com a aquisição desses dados é verificar se há efeito de práticas de manejo (linha e entre linha) na retenção de água do solo bem como verificar a extensão de tais efeitos ao longo do perfil do solo. Segundo (SERAFIM et al., 2011), essas lavouras de café pertencem a um sistema intensivo de cultivo denominado de Sistema AP Romero que é caracterizado pela aplicação de altas doses de gesso agrícola na linha de cultivo do café e pelo cultivo de braquiária (*Brachiaria* sp.) nas entre linhas. Em combinação, essas medidas visam melhorar a capacidade do solo em fornecer água para a cultura do café de sequeiro (não irrigado) dadas as limitações de ordem climática e pedológica da região do Alto do São Francisco.

Para determinações do conteúdo de água sob as tensões aplicadas, as amos-

tras foram devidamente preparadas e saturadas por meio de elevação gradual de uma lâmina de água, em bandeja. Quando saturadas elas foram submetidas às seguintes tensões: 1, 2, 4, 6 e 10 kPa na mesa de tensão; 33, 100, 500 e 1500 kPa na membrana extratora de Richards (CLAESSEN et al., 1997). Após atingir o equilíbrio hídrico em cada tensão, as amostras foram pesadas e então submetidas à tensão seguinte, constituindo o método por secamento. Após a última tensão, as amostras foram secas em estufa a ± 105 °C, por 24 horas, para determinação do conteúdo de água.

4.1 Medidas de curvatura

Os procedimentos de inferência em modelos de regressão não linear fazem uso de argumentos assintóticos e frequente analogia com modelos de regressão linear (DRAPER; SMITH, 1998; SEBER; WILD, 2003). Em outras palavras, as inferências são baseadas numa aproximação linear do modelo não linear.

Para avaliar a qualidade da aproximação linear, Bates e Watts (1980) propuseram medidas de curvatura que se baseiam na curvatura da função não linear $f(x, \theta)$ por meio da magnitude da segunda derivada $\partial^2 f / \partial \theta_j \partial \theta_k$, com $j, k = 1, \dots, p$, p é a dimensão de θ , n é o número de observações. Se o valor absoluto da segunda derivada é pequeno, então o modelo é aproximadamente linear e as suposições assintóticas se tornam aceitáveis. Seja $\mathbf{u} = (u_1, \dots, u_n)^\top$ um vetor que representa uma direção de interesse no espaço vetorial a partir de $\hat{\theta}$ e defina

$$A_{jk} = \sum_{i=1}^n u_i \frac{\partial^2 f(\mathbf{x}_i, \hat{\theta})}{\partial \theta_j \partial \theta_k}. \quad (4.1)$$

A matriz $\mathbf{A}_{(p \times p)} = \{A_{jk}\}$ representa o valor da segunda derivada na direção \mathbf{u} . Para obter medidas de curvatura é necessário padronizar as derivadas pelo valor do vetor gradiente na mesma direção. Seja $\mathbf{X} = \mathbf{QR}$ a decomposição QR da matriz

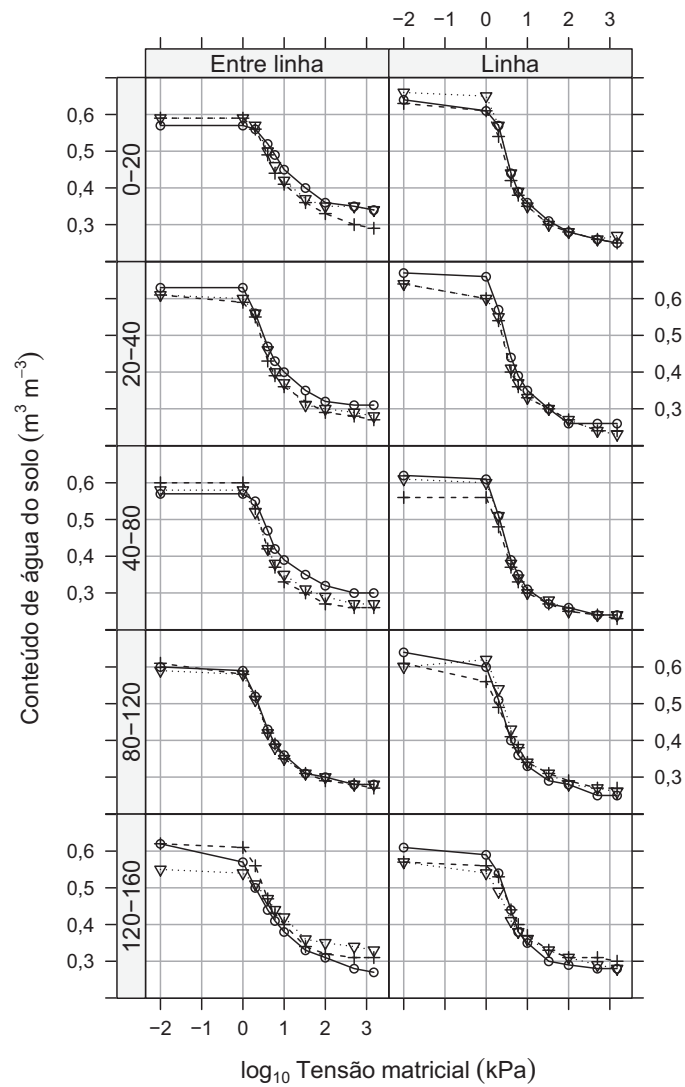


Figura 5 Conteúdo de água do solo ($\text{m}^3 \text{m}^{-3}$) em função do \log_{10} da tensão matricial (kPa) organizado pelas combinações entre posição (nas colunas) e profundidade de coleta (nas linhas). As três unidades experimentais em cada combinação foram identificadas pelos tipos de pontos e linhas

$$\mathbf{X}_{(n \times p)} = \left. \frac{\partial f(\mathbf{x}_i, \hat{\boldsymbol{\theta}})}{\partial \theta_j} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}, \text{ e seja}$$

$$\mathbf{C}_{(p \times p)} = (\mathbf{R}^{-1})^\top \mathbf{A} \mathbf{R}^{-1}. \quad (4.2)$$

Bates e Watts (1980) usam \mathbf{C} para definirem as medidas de curvatura relativas à não linearidade. Se \mathbf{u} pertence ao espaço vetorial gerado pelas colunas de \mathbf{X} ($\mathbf{u} = \mathbf{X}\boldsymbol{\theta}$ para um dado $\boldsymbol{\theta}$), então \mathbf{C} representa a curvatura devido ao efeito de parâmetros. Se $\mathbf{X}^\top \mathbf{u} = \mathbf{0}$ então \mathbf{C} representa a curvatura intrínseca do modelo. Pode-se mostrar que a curvatura intrínseca do modelo depende apenas da forma da superfície esperada, gerada pelos valores preditos pelo modelo $f(\mathbf{x}, \boldsymbol{\theta})$, à medida que $\boldsymbol{\theta}$ varia e que por isso é invariante em relação à parametrização adotada (SEBER; WILD, 2003).

Para avaliar a qualidade inferencial da parametrização proposta, as medidas de curvatura (intrínseca e devido ao efeito de parâmetros) de Bates e Watts (1980) foram calculadas para o modelo Genuchten (1980) ajustado sob a parametrização considerada por Omuto (2007), ou seja,

$$U(x) = U_r + \frac{U_s - U_r}{(1 + (\exp\{a\} \exp\{x\})^n)^{1-1/n}}, \quad (4.3)$$

e a desenvolvida para inferência direta sobre os parâmetros I e S (3.7). Então foi feito o ajuste do modelo em cada parametrização, (3.7) e (4.3), para cada uma das 30 unidades amostrais. Os ajustes individuais são um passo importante pois servem de análise exploratória, da qual se pode ter informações sobre a qualidade de ajuste do modelo por medidas como o coeficiente de determinação, avaliar as suposições como homocedasticidade. Gráficos de dispersão das estimativas contra valores das variáveis independentes, são úteis para fornecer valores iniciais para estimação do modelo não linear de efeitos mistos.

4.2 Análises em dois estágios

As estimativas dos parâmetros S e I para cada unidade experimental podem ser obtidas por meio de (2.5) e (2.6) após o ajuste do modelo original (4.3) ou diretamente pelo ajuste do modelo reparametrizado (3.7). As 30 estimativas de cada parâmetro foram então consideradas variáveis dependentes em uma análise de variância em que o modelo linear correspondente descreve um experimento fatorial 2×5 . Os efeitos dos fatores posição e profundidade então são testados para cada parâmetro da CRA por meio da análise de variância. Para efeito de comparação, uma análise de variância ponderada foi também considerada, na qual os pesos são as variâncias das estimativas dos parâmetros. Para I e S elas são obtidas pelo método delta.

A análise baseada apenas nas estimativas foi denominada análise em dois estágios simples (ADES) e a outra de análise em dois estágios ponderada (ADEP) por considerar as variâncias.

4.3 Modelo não linear de efeitos mistos (NLMEM)

O modelo não linear de efeitos mistos (NLMEM) para a j -ésima observação na i -ésima unidade experimental é representado por

$$y_{ij} \sim N(f(\boldsymbol{\theta}_i, \mathbf{x}_{ij}), \sigma^2), \quad (4.4)$$

em que y_{ij} é a resposta observada, \mathbf{x}_{ij} é um vetor de variáveis independentes, f é uma função não linear do vetor de parâmetros $\boldsymbol{\theta}_i$, de tamanho p , σ^2 é a variância residual. Em geral não se antecipam restrições ao domínio do vetor de variáveis independentes. Os parâmetros podem mudar de indivíduo para indivíduo de acordo com grupos e covariáveis descritos por \mathbf{x}_{ij} , o que é representado por mais um nível

de especificação de modelo sobre o termo θ_i ,

$$\theta_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, \quad (4.5)$$

$$\mathbf{b}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{D}), \quad (4.6)$$

em que $\boldsymbol{\beta}$ é um vetor de dimensão maior ou igual à p associado à parâmetros de fatores de efeitos fixos, \mathbf{b}_i é um vetor de dimensão q que representa efeitos aleatórios associados à unidade experimental i . As matrizes \mathbf{A}_i e \mathbf{B}_i são matrizes de delineamento de dimensão apropriada e \mathbf{D} é a matriz de covariância dos efeitos aleatórios. Por meio da correta especificação das matrizes \mathbf{A}_i e \mathbf{B}_i é que se acomoda os efeitos de variáveis independentes, sejam elas tratamentos, covariáveis ou níveis de amostragem.

Considerando os dados descritos e a parametrização do modelo (3.7) desenvolvido para estes, o NLMEM e seus termos, ao nível da unidade experimental i , têm a seguinte estrutura

$$f(\boldsymbol{\theta}_i, \mathbf{x}_{ij}) = \theta_{1i} - \frac{\theta_{2i}}{\theta_{3i}} \frac{\left(\frac{2\theta_{3i}-1}{\theta_{3i}-1}\right)^{2-1/\theta_{3i}}}{\left(1 + \left(\frac{\theta_{3i}}{\theta_{3i}-1}\right) \exp\{\theta_{3i}(x_{ij} - \theta_{4i})\}\right)^{1-1/\theta_{3i}}}, \quad (4.7)$$

$$\boldsymbol{\theta}_i = \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \\ \theta_{3i} \\ \theta_{4i} \end{bmatrix} = \begin{bmatrix} U_r \\ S \\ n \\ I \end{bmatrix} + \begin{bmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \\ b_{4i} \end{bmatrix} = \boldsymbol{\beta} + \mathbf{b}_i, \quad \mathbf{b}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{D}). \quad (4.8)$$

As matrizes matrizes \mathbf{A}_i e \mathbf{B}_i , por sua vez, são dependentes do processo de construção de modelos. Ao invés de escrever as matrizes é mais cômodo representá-las por meio das fórmulas dos modelos que as definem conforme notação apresentada por Wilkinson e Rogers (1973). Seja θ_k , $k = 1, \dots, 4$, um dos parâmetros,

a sequência de fórmulas definem porções das matrizes \mathbf{A}_i para efeitos fixos e \mathbf{B}_i para os efeitos aleatórios com relação à θ_k baseadas nas variáveis independentes ao nível de indivíduo

$$\mathbf{A}_{i(k)} : \begin{cases} \theta_k \sim \text{PF} + \text{PS} + \text{PF} \cdot \text{PS} \\ \theta_k \sim \text{PF} + \text{PS} \\ \theta_k \sim \text{PF} \\ \theta_k \sim 1 \end{cases} \quad (4.9)$$

$$\mathbf{B}_{i(k)} : \begin{cases} \theta_k \sim \text{UE} \\ \theta_k \sim 1 \end{cases} \quad (4.10)$$

A notação $\theta_k \sim \text{PF} + \text{PS} + \text{PF} \cdot \text{PS}$ representa que o parâmetro θ_k tem valores dependente do efeito de profundidade (PF), posição de amostragem (PS) e sua interação (PF · PS). A notação $\theta_k \sim \text{UE}$ indica que o valores de θ_k dependem da unidade experimental (UE). Por fim, $\theta_k \sim 1$ diz que o valor de θ_k é uma constante.

Nessas especificações, o modelo simplifica da primeira para última linha de forma que o modelo seguinte é sempre um caso particular do modelo anterior, ou seja, um modelo aninhado. Verifica-se que um grande número de modelos intermediários podem existir considerando as combinações geradas entre as partes de efeitos fixos e aleatórios e os quatro parâmetros do modelo. No entanto, as especificações devem preconizar os parâmetros U_r , S e I uma vez que n é um parâmetro de maior ordem desprovido de interpretação prática.

Para comparar as especificações do modelo foi feito o abandono sequencial de termos a partir do modelo maior até o modelo sem considerar qualquer efeito. O abandono de termos de efeito fixo foi avaliado pelos testes F e t marginais de Wald (PINHEIRO; BATES, 2000). Quando realizado o abandono de

termos de efeito aleatório considerou-se o teste da razão de verossimilhanças.

5 RESULTADOS E DISCUSSÃO

Estimativas dos parâmetros foram obtidas para as 30 unidades experimentais sob as duas parametrizações do modelo van Genuchten. Os valores de R^2 apontaram boa medida de ajuste pois ficaram entre 98,96% e 99,88%. Pelas estimativas intervalares, considerando os seis parâmetros definidos pelas duas parametrizações, U_s , U_r , a , n , S e I , observamos variação entre unidades experimentais dentro de uma mesma cela experimental (combinação de posição e profundidade), entre níveis de profundidade e entre posições de amostragem (Figura 6). Verifica-se que os parâmetros U_r e n , por serem comuns as duas parametrizações, tiveram mesmas estimativas intervalares independentemente da parametrização.

Ainda pela Figura 6 pode-se antecipar algum efeito da profundidade em S , U_s e U_r uma vez que os valores pontuais se dispõem em forma de arco em relação à profundidade, típico de uma função de segundo grau. O parâmetro I apresenta uma relação mais linear com a profundidade. Por sua vez, o parâmetro n mostra-se pouco dependente, considerando a amplitude dos intervalos de confiança em relação ao efeito de profundidade e posição de amostragem.

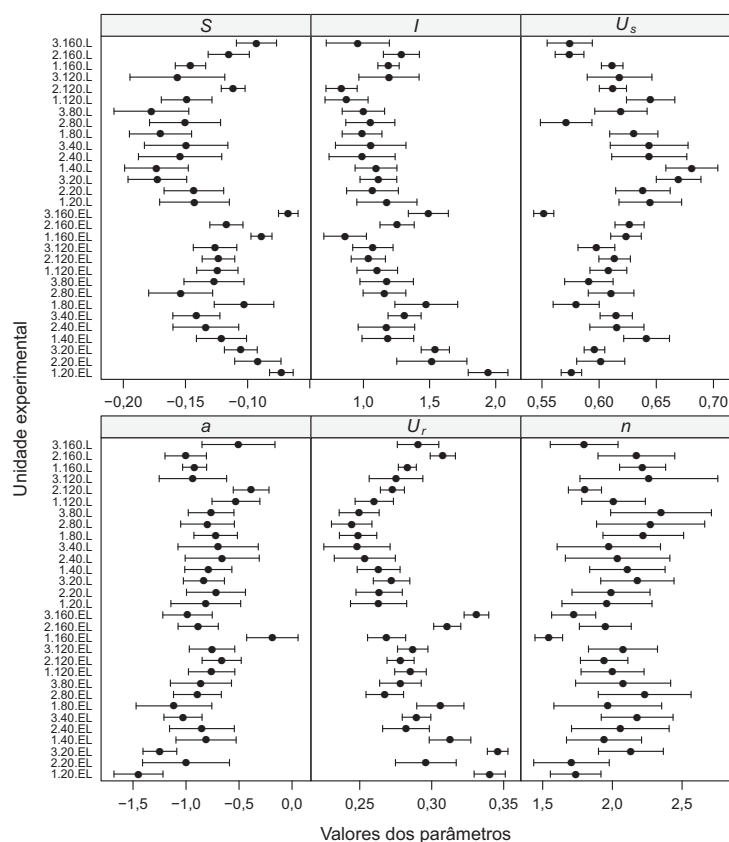


Figura 6 Estimativas intervalares (95%) pelo método de Wald para os parâmetros da CRA considerando as duas parametrizações do modelo van Genuchten. Os rótulos no eixo das coordenadas representam o índice da repetição, o nível de profundidade e o nível de posição, separados por ponto

O ajuste individual foi realizado para cálculo das curvaturas em cada reparametrização. O conhecimento das medidas de curvatura não é obrigatório e nem eliminatório mas classificatório na escolha de uma parametrização, caso existam várias opções. Considerando o total de 30 curvas ajustadas, em apenas cinco delas a curvatura devido a parametrização do modelo reparametrizado foi superior ao

modelo original, ou seja, em 5/6 das curvas o modelo reparametrizado apontou curvatura menor (Figura 7). A curvatura intrínseca por sua vez não se alterou com a parametrização, como já é de conhecimento. Além disso, verificou-se que as curvaturas são positivamente relacionadas.

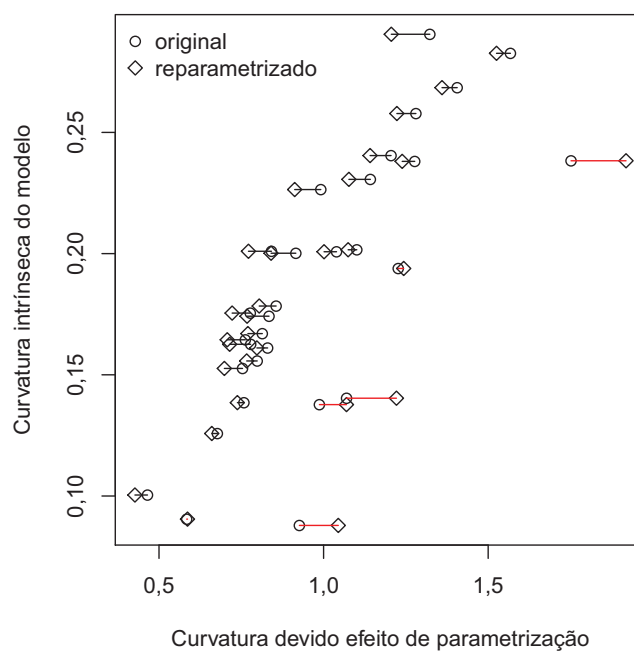


Figura 7 Curvatura intrínseca e devido ao efeito de parâmetros para as parametrizações do modelo van Genuchten. Os segmentos ligam às curvaturas obtidas para as parametrizações do modelo ajustado aos dados da mesma unidade experimental

Nas análises em dois estágios, as 30 estimativas para cada um dos parâmetros foram consideradas variáveis dependentes em um modelo linear que representa o experimento fatorial 2×5 , com 3 repetições. O resumo do quadro de análise de variância para a ADES indicou não haver interação entre profundidade e posição de amostragem para nenhum dos parâmetros (Tabela 1), ao passo que indicou

efeito da posição de amostragem em todos os parâmetros, e efeito de profundidade para todos exceto para a e n , ao nível de 5%. Pela ADEP, houve interação para os parâmetros U_r , a e n , e não houve efeito da profundidade em U_s (Tabela 2).

A ADES desconsidera em grande parte as informações referentes ao ajuste do modelo não linear. Embora a ADEP considere as informações com relação ao nível de precisão das estimativas por meio dos pesos, a informação sobre as covariâncias entre as estimativas ainda não são aproveitadas. Apenas por essas duas simples e evidentes desvantagens das análises em dois estágios considera-se que o NLMEM seja vantajoso uma vez que os parâmetros serão estimados pela especificação de um único modelo, o que contempla as covariâncias entre todas as estimativas além de acomodar o efeito da variação entre unidades experimentais da variação entre observações dentro da mesma unidade experimental.

Sob o enfoque dos NLMEM, todos os parâmetros nas duas parametrizações do modelo van Genuchten foram especificados como funções lineares dos efeitos principais e interação envolvendo os fatores profundidade e posição de amostragem. Inicialmente o termo de efeito aleatório foi especificado também para todos os parâmetros para representar o efeito de unidade experimental. Porém, o modelo dessa forma ficou demasiado complexo e não se obteve a convergência. Então se simplificou a especificação dos efeitos aleatórios com ênfase nos parâmetros que ocorrem de forma linear no modelo, no caso U_r , U_s e S , e a convergência foi alcançada.

Na figura 8 estão os valores observados e preditos pelo modelo van Genuchten reparametrizado considerando os efeitos fixos de profundidade e posição de amostragem ($\sim PF + PS + PF \cdot PS$) para todos os parâmetros e efeito aleatório de unidade experimental para U_r e S . Os valores preditos não diferem daqueles obtidos com o modelo original com efeito aleatório em U_r e U_s . No entanto,

Tabela 1 Resumo do quadro de análise de variância considerando as estimativas dos parâmetros como variável dependente em uma análise em dois estágios simples (ADES)

FV	DF	Ur	Us	a	n	I	S
PF	4	4,585 (<,01) ^a	4,000 (0,015)	2,227 (0,103)	2,726 (0,058)	4,146 (0,013)	5,240 (<,01)
PS	1	28,319 (<,01)	7,210 (0,014)	4,455 (0,048)	5,570 (0,029)	14,463 (<,01)	23,974 (<,01)
PF:PS	4	2,375 (0,087)	2,185 (0,108)	1,417 (0,265)	1,115 (0,377)	2,080 (0,122)	1,391 (0,273)
Resíduo	20	0,000275 ^b	0,000501	0,0438	0,0263	0,0271	0,000358

^avalor de F (p valor), ^bquadrado médio.

Tabela 2 Resumo do quadro de análise de variância considerando as estimativas dos parâmetros como variável dependente em uma análise em dois estágios ponderada (ADEP)

FV	DF	U_r	U_s	a	n	l	S
PF	4	11,115 (<,01) ^a	2,055 (0,125)	4,811 (<,01)	4,334 (0,011)	7,460 (<,01)	6,373 (<,01)
PS	1	30,857 (<,01)	5,375 (0,031)	4,747 (0,041)	7,125 (0,015)	14,451 (<,01)	16,592 (<,01)
PF.PS	4	3,261 (0,033)	1,215 (0,336)	2,917 (0,047)	3,531 (0,025)	2,580 (0,069)	2,418 (0,082)
Resíduo	20	6,466 ^b	10,52	2,543	2,102	3,643	5,881

^avalor de F (p valor), ^bquadrado médio.

os valores de log-verossimilhança calculados, 853,20 e 852,98, apontaram que o modelo reparametrizado se ajustou melhor aos dados.

Calculando o R^2 para cada unidade experimental, teve-se que os valores ficaram entre 98,02% e 99,74%, sempre abaixo mas muito próximos dos valores observados para o ajuste das curvas individuais que não têm qualquer imposição de estrutura experimental. A maior diferença de R^2 foi de 1,27% entre os ajustes individuais e o NLMEM, ou seja, uma diferença que sugere não haver efeito aleatório de unidade experimental nos parâmetros I e n .

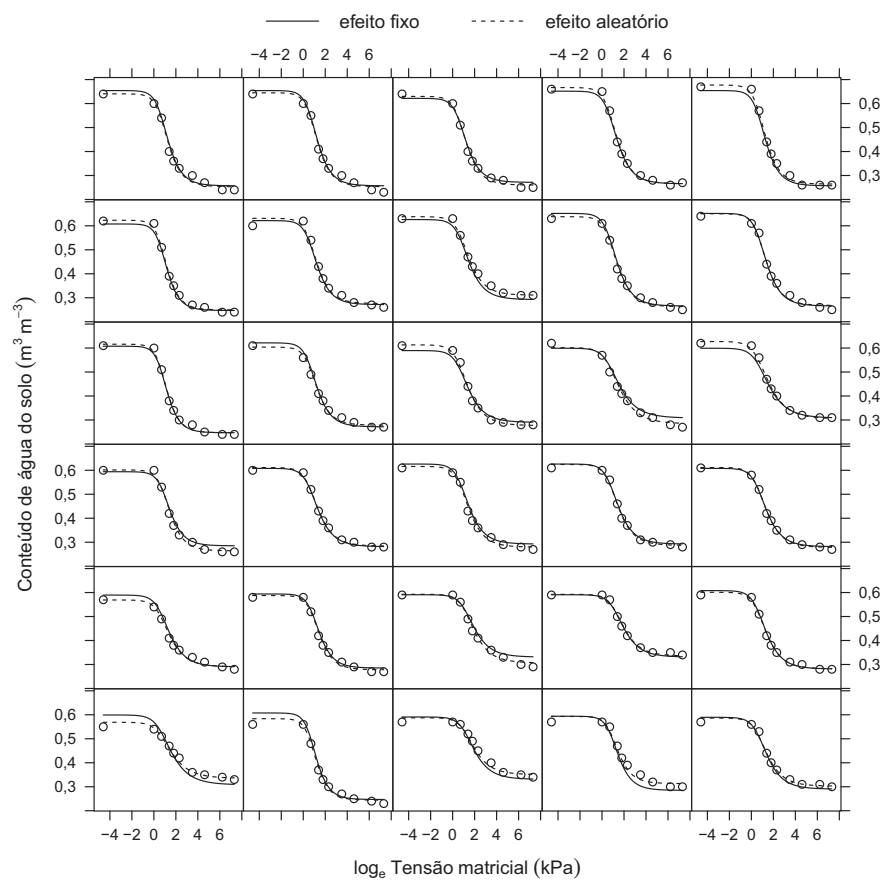


Figura 8 Valores observados e preditos considerando os efeitos fixos de posição e profundidade de amostragem e o efeito aleatório da unidade experimental pelo modelo van Genuchten reparametrizado

Os testes sobre o efeito dos fatores experimentais indicam haver interação apenas para o parâmetro U_s do modelo original (Tabela 3) enquanto que o modelo reparametrizado indicou interação entre os fatores para o parâmetro S e I (Tabela 4). Tanto as diferenças nos valores de log-verossimilhança quanto as diferenças nos valores dos testes de Wald podem estar relacionadas às medidas de curvatura (Figura 7) que favoreceram o modelo reparametrizado, indicando que a verossi-

milhança para este pode ser uma função mais quadrática em torno das estimativas dos parâmetros.

Tabela 3 Valores de F para o teste de Wald marginal para o efeito de profundidade e posição de amostragem nos parâmetros do modelo van Genuchtem original

parâmetro	efeito	GL	completo	reduzido
U_r	intercepto	1	1055,23 (<,001) ^a	1203,56 (<,001)
	PF	4	3,54 (0,008)	4,66 (0,001)
	PS	1	20,09 (<,001)	24,70 (<,001)
	PF·PS	4	2,35 (0,055)	2,73 (0,030)
U_s	intercepto	1	2764,91 (<,001)	2898,62 (<,001)
	PF	4	1,27 (0,281)	1,60 (0,175)
	PS	1	13,24 (<,001)	14,87 (<,001)
	PF·PS	4	2,75 (0,029)	2,69 (0,032)
a	intercepto	1	93,87 (<,001)	169,68 (<,001)
	PF	4	3,21 (0,014)	6,26 (<,001)
	PS	1	7,60 (0,006)	22,45 (<,001)
	PF·PS	4	2,01 (0,094)	5,68 (<,001)
n	intercepto	1	292,72 (<,001)	606,42 (<,001)
	PF	4	3,22 (0,014)	2,94 (0,021)
	PS	1	2,81 (0,095)	6,19 (0,014)
	PF·PS	4	1,31 (0,266)	
	II		852,977 (231) ^b	850,072 (235)

^avalor de F (p valor), ^bvalor de log-verossilhança (graus de liberdade).

A partir dos testes de Wald, uma simplificação foi abandonar o termo PF · PS para o parâmetro n . Em ambas parametrizações, o termo abandonado foi não significativo pelo teste de razão de verossimilhanças. Com a simplificação do modelo houve modificação dos valores da estatística F que apontou efeito da interação entre os fatores para os demais parâmetros em qualquer parametrização. Tal resultado evidencia que os teste sobre os efeitos dentro de um NLMEM considera a informação das covariâncias entre as estimativas dos parâmetros.

Tabela 4 Valores de F para o teste de Wald marginal para o efeito de profundidade e posição de amostragem nos parâmetros do modelo van Genuchten reparametrizado

parâmetro	efeito	GL	completo	reduzido
U_r	intercepto	1	1054,29 (<,001) ^a	1196,81 (<,001)
	PF	4	3,53 (0,008)	4,62 (0,001)
	PS	1	20,27 (<,001)	24,74 (<,001)
	PF·PS	4	2,37 (0,054)	2,72 (0,030)
S	intercepto	1	120,01 (<,001)	173,43 (<,001)
	PF	4	7,22 (<,001)	6,84 (<,001)
	PS	1	29,22 (<,001)	43,96 (<,001)
	PF·PS	4	2,61 (0,036)	4,59 (0,001)
I	intercepto	1	419,10 (<,001)	506,77 (<,001)
	PF	4	7,96 (<,001)	10,20 (<,001)
	PS	1	28,52 (<,001)	38,49 (<,001)
	PF·PS	4	4,01 (0,004)	5,71 (<,001)
n	intercepto	1	297,11 (<,001)	622,86 (<,001)
	PF	4	3,86 (0,005)	3,46 (0,009)
	PS	1	2,51 (0,115)	5,55 (0,019)
	PF·PS	4	1,37 (0,246)	
	ll		853,202 (231) ^b	850,159 (235)

^avalor de F (p valor), ^bvalor de log-verossilhança (graus de liberdade).

O desvio-padrão do efeito aleatório associado à U_s é superior ao desvio-padrão associado à U_r , o que significa que as diferenças entre as unidades experimentais tendem a ser mais relacionadas às características ligadas aos valores de umidade em baixas tensões, como a estrutura do solo (Tabela 5). Essa comparação só é possível porque tais parâmetros têm mesma unidade de medida. A correlação negativa entre os efeitos aleatórios indica que uma unidade experimental de maior U_r tenderá a ter menor U_s , no entanto essa correlação não foi diferente de zero pelo teste de razão de verossimilhanças (valor $p = 0,31$) e por isso pode considerar-se que esses efeitos aleatórios são independentes. Da mesma forma, a correlação

positiva indicou que uma unidade experimental de maior U_r tenderá a ter maior S , e essa correlação foi diferente de zero pelo teste de razão de verossimilhanças (valor $p \leq 0,0001$) (Tabela 6). Tais resultados podem indicar que U_s e U_r são parâmetros ligados à características do solo pouco correlacionadas ao passo que U_r e S são dependentes de características comuns.

Tabela 5 Estimativas dos desvios-padrões dos termos de efeito aleatório para o modelo van Genuchten original

Parâmetro	Completo		Reduzido	
	desvio-padrão	correlação	desvio-padrão	correlação
U_r	0,01359		0,01347	
U_s	0,01546	-0,212	0,01540	-0,206
Resíduo	0,01168		0,01183	

Tabela 6 Estimativas dos desvios-padrões dos termos de efeito aleatório para o modelo van Genuchten reparametrizado

Parâmetro	Completo		Reduzido	
	desvio-padrão	correlação	desvio-padrão	correlação
U_r	0,01368		0,01356	
S	0,00815	0,726	0,00823	0,729
Resíduo	0,01169		0,01184	

Os métodos de análise, ADES, ADEP e NLMEM, forneceram estimativas pontuais e intervalares distintas (Figura 9) para os parâmetros. Ao considerar as covariâncias entre as estimativas e o efeito aleatório sobre alguns parâmetros, os valores pontuais para o NLMEM foram ajustados para essas quantidades. Os intervalos de confiança para a ADES têm a mesma amplitude uma vez que esta é função apenas do número de repetições. Para a ADEP, os intervalos têm amplitude variável devido uso de pesos, que são as variâncias das estimativas, provenientes

do ajuste do modelo não linear para cada unidade experimental. No NLMEM, o intervalo de confiança é função da matriz de informação de Fisher que depende de todos os parâmetros presentes no modelo. Dentre os três procedimentos, o NLMEM foi o que apresentou, em média, intervalos mais estreitos, ou seja, conferiu maior precisão às estimativas dos parâmetros. Em algumas situações, pode-se verificar que as estimativas pelo NLMEM foram distantes das fornecidas pela ADES ou ADEP e que os intervalos de confiança foram disjuntos.

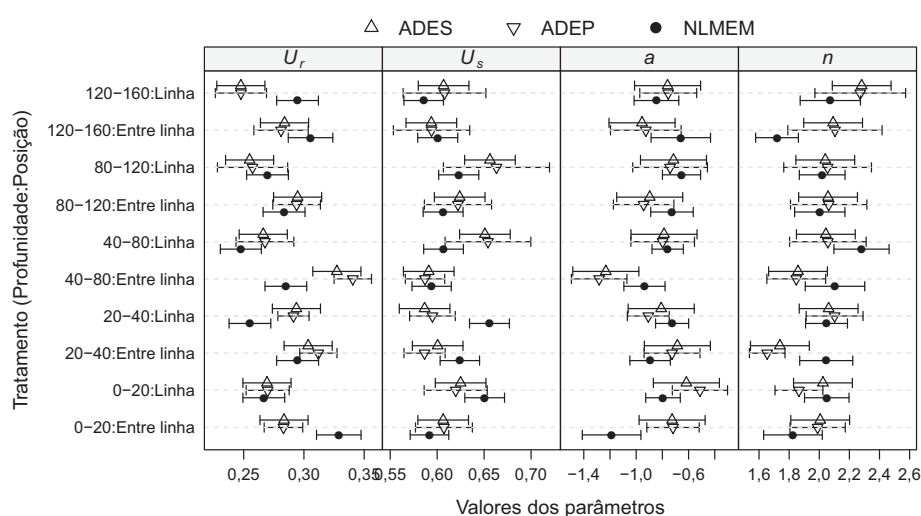


Figura 9 Estimativas intervalares (95%) para os parâmetros do modelo van Genuchten original considerando a análise pelo modelo não linear de efeitos mistos (NLMEM) e os resultados da análise em dois estágios simples (ADES) e ponderada (ADEP)

Na figura 10 tem-se conclusões semelhantes às aquelas obtidas a partir figura 9 sobre os procedimentos de análise. Nesse ponto porém, obteve-se ainda as estimativas intervalares dos parâmetros I e S por meio do método delta aplicado no NMLEM com a parametrização original. Tais cálculos envolveram (2.5) e (2.6) e suas derivadas de primeira ordem, o vetor de estimativas dos parâmetros

$\hat{\theta}$, de dimensão 4 parâmetros \times 2 posições \times 5 profundidades = 40, e a matriz de variância-covariância $\hat{\Sigma}_{\theta}$ de dimensão 40 \times 40. Antes de aplicar o método, deve-se reconhecer que $\hat{\theta}$ e $\hat{\Sigma}_{\theta}$ não representam diretamente as estimativas para cada tratamento, mas um conjunto de contrastes, relativos a uma particular restrição paramétrica linear, como a soma dos efeitos ser zero. Operações matriciais precisam ser feitas para se obter, a partir dos contrastes, as estimativas e suas variâncias para cada tratamento, de forma a ser facilitada a aplicação do método delta. Apesar de serem operações lineares, a dimensão das matrizes envolvidas e por vezes o tipo de contraste considerado, pode tornar inviável tal procedimento.

O método delta retornou estimativas pontuais muito próximas, sendo a maior diferença de 2,3% para S e 1,4% para I ao comparar com os valores dados pelo NLMEM na versão reparametrizada do van Genuchten. No entanto, a amplitude dos intervalos de confiança pelo método delta foram maiores, em um dos casos, na profundidade 0-20cm e linha (0-20:Linha), amplos demais para serem úteis na prática.

De forma geral, o que se observa é que os intervalos de confiança pelo NLMEM são mais estreitos, conferindo maior precisão às estimativas, o que está ligado à um maior poder para testar hipóteses sobre os parâmetros. Por exemplo, a hipótese nula: o valor de S é igual nas duas posições de amostragem para a camada 0-20 cm, foi rejeitada pelo NLMEM (Figura 10), dada grande disjunção dos intervalos de confiança, e aceita pelo método delta e análise em dois estágios, visto que os intervalos se sobrepuseram. Nesse ponto, o MNLEM apresenta resultados mais aceitáveis do ponto de vista prático pois espera-se que haja diferenças entre os manejos na linha e entre linha na primeira camada do solo.

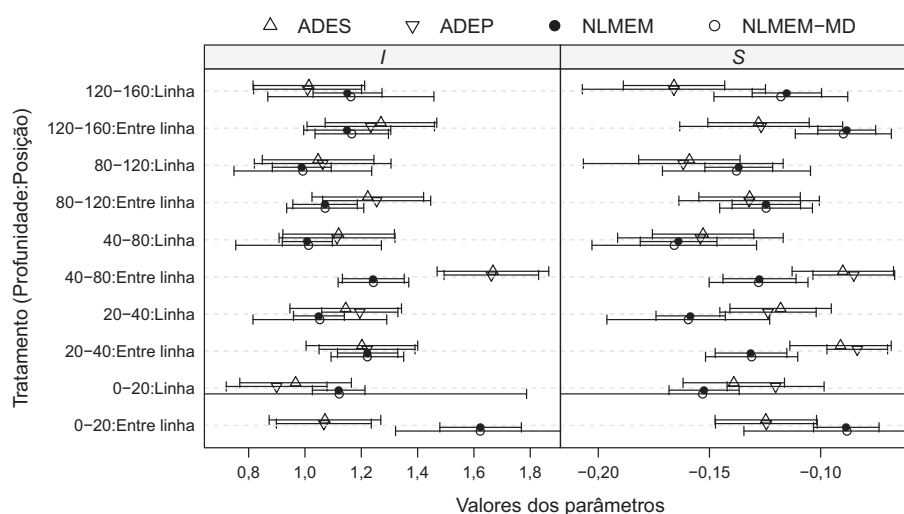


Figura 10 Estimativas intervalares (95%) para os parâmetros I e S considerando os resultados da análise em dois estágios simples (ADES) e ponderada (ADEP), o modelo não linear de efeitos mistos para a versão reparametrizada do van Genuchten (NLMEM) e o método delta sobre o modelo não linear de efeitos mistos na parametrização original (NLMEM-MD)

A CRA predita pelo NLMEM, independentemente da parametrização, mostra que as diferenças devido à posição de amostragem são menores à medida que aumenta a profundidade (Figura 11). Na primeira camada a CRA para linha tem uma amplitude maior ($U_s - U_r$), ou um valor de S maior, se comparado à da entre linha. Na última camada verifica-se uma diferença praticamente nula uma vez que os valores preditos para um nível de posição estão dentro das bandas de confiança do outro praticamente em todo domínio de tensão avaliado.

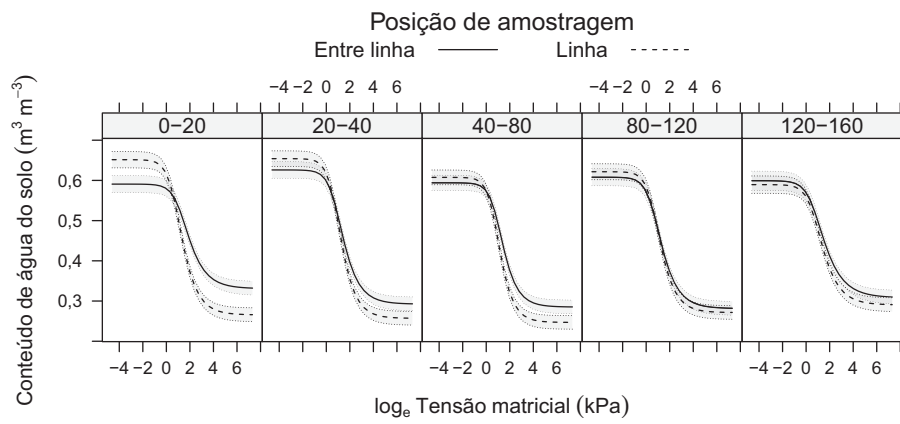


Figura 11 Estimativas intervalares (95%) para a curva de retenção de água do solo em função da profundidade e posição de amostragem do solo

6 CONCLUSÕES

Conclui-se que o modelo não linear de efeitos mistos é um método de análise mais adequado para representar os dados uma vez que toda informação está contida em um único modelo que permite acomodar os efeitos de termos fixos e aleatórios, comparar modelos e fazer predições para a CRA.

REFERÊNCIAS

- ANDRADE, R. S.; STONE, L. F. Índice S como indicador da qualidade física de solos do cerrado brasileiro. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 13, n. 4, p. 382–388, Aug. 2009.
- BATES, D.; WATTS, D. Relative curvature measures of nonlinearity. **Journal of the Royal Statistical Society. Series B (Methodological)**, London, v. 42, n. 1, p. 1–25, Jan. 1980.
- BRADY, N. C.; WEIL, R. R. **Elements of the nature and properties of soils**. 3 ed. [S.l.]: Prentice Hall, 2009. 624 p.
- CALAMA, R.; MONTERO, G. Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. **Canadian Journal of Forest Research**, Ottawa, v. 34, n. 1, p. 150–163, Jan. 2004.
- CALEGARIO, N. et al. Modeling dominant height growth based on nonlinear mixed-effects model: a clonal Eucalyptus plantation case study. **Forest Ecology and Management**, Amsterdam, v. 204, n. 1, p. 11–21, Jan. 2005.
- CARDUCCI, C. E. et al. Modelagem da curva de retenção de água de Latossolos utilizando a equação duplo van Genuchten. **Revista Brasileira de Ciência do Solo**, Campinas, v. 35, n. 1, p. 77–86, Feb. 2011.
- CHESWORTH, W. (Ed.). **Encyclopedia of Soil Science (Encyclopedia of Earth Sciences Series)**. [S.l.]: Springer, 2007. 902 p.
- CLAESSEN, M. E. C.; BARRETO, W. O.; PAULA, J. L. **Manual de métodos de análises do solo**. 2. ed. Rio de Janeiro: Embrapa, 1997. 212 p.
- DEXTER, A. Soil physical quality: Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. **Geoderma: an international journal of soil science**, Amsterdam, v. 120, n. 3-4, p. 201–214, June 2004.
- DEXTER, A. Soil physical quality: Part II. Friability, tillage, tilth and hard-setting. **Geoderma: an international journal of soil science**, Amsterdam, v. 120, n. 3-4, p. 215–225, June 2004.
- DEXTER, A. Soil physical quality: Part III: Unsaturated hydraulic conductivity and general conclusions about S-theory. **Geoderma: an international journal of soil science**, Amsterdam, v. 120, n. 3-4, p. 227–239, June 2004.

DEXTER, A.; BIRD, N. Methods for predicting the optimum and the range of soil water contents for tillage based on the water retention curve. **Soil and Tillage Research**, Amsterdam, v. 57, n. 4, p. 203–212, Jan. 2001.

DRAPER, N.; SMITH, H. **Applied regression analysis**. 3. ed. New York: Wiley-Interscience, 1998. 706 p.

GENUCHTEN, M. T. van. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. **Soil Science Society of America Journal**, Madison, v. 44, n. 5, p. 892–898, May 1980.

GREENLAND, S. Principles of multilevel modelling. **International Journal of Epidemiology**, Oxford, v. 29, n. 1, p. 158–167, Feb. 2000.

JORDAN, L. et al. Multilevel nonlinear mixed-effects models for the modeling of earlywood and latewood microfibril angle. **Forest Science**, Lawrence, v. 51, n. 4, p. 357–371, Aug. 2005.

MACHADO, J. L. et al. Inter-relações entre as propriedades físicas e os coeficientes da curva de retenção de água de um latossolo sob diferentes sistemas de uso. **Revista Brasileira de Ciência do Solo**, Campinas, v. 32, n. 2, p. 495–502, Apr. 2008.

MAIA, C. E. Índice S para avaliação da qualidade física de solos. **Revista Brasileira de Ciência do Solo**, Campinas, v. 35, n. 6, p. 1959–1965, Dec. 2011.

MENG, S. X.; HUANG, S. Incorporating correlated error structures into mixed forest growth models: prediction and inference implications. **Canadian Journal of Forest Research**, Ottawa, v. 40, n. 5, p. 977–990, May 2010.

MUALEM, Y. A new model for predicting the hydraulic conductivity of unsaturated porous media. **Water Resources Research**, Washington, v. 12, n. 3, p. 513, June 1976.

NIKOLAIDIS, N. et al. Non-linear response of a mixed land use watershed to nitrogen loading. **Agriculture, Ecosystems & Environment**, Amsterdam, v. 67, n. 2-3, p. 251–265, Feb. 1998.

NOTHDURFT, A.; KUBLIN, E.; LAPPI, J. A non-linear hierarchical mixed model to describe tree height growth. **European Journal of Forest Research**, v. 125, n. 3, p. 281–289, Mar. 2006.

OMUTO, C. T. **HydroMe**: estimation of soil hydraulic parameters from experimental data. [S.l.]: R Development Core Team, 2007. Software

- OMUTO, T. et al. Nonlinear mixed effect modelling for improved estimation of water retention and infiltration parameters. **Journal of Hydrology**, Amsterdam, v. 330, n. 3-4, p. 748–758, Nov. 2006.
- PINHEIRO, J.; BATES, D. **Mixed-effects models in S and S-plus**. 1. ed. New York: Springer, 2000. 548 p.
- QUENÉ, H.; BERGH, H. van den. On multi-level modeling of data from repeated measures designs: a tutorial. **Speech Communication**, Amsterdam, v. 43, n. 1-2, p. 103–121, June 2004.
- REKAYA, R.; WEIGEL, K.; GIANOLA, D. Hierarchical nonlinear model for persistency of milk yield in the first three lactations of Holsteins. **Livestock Production Science**, Amsterdam, v. 68, n. 2-3, p. 181–187, Mar. 2001.
- REYNOLDS, W. et al. Indicators of good soil physical quality: density and storage parameters. **Geoderma: an international journal of soil science**, Amsterdam, v. 110, n. 1-2, p. 131–146, Nov. 2002.
- REYNOLDS, W. et al. Use of indicators and pore volume-function characteristics to quantify soil physical quality. **Geoderma: an international journal of soil science**, Amsterdam, v. 152, n. 3-4, p. 252–263, Sept. 2009.
- SCHABENBERGER, O.; PIERCE, F. J. **Contemporary statistical models for the plant and soil Sciences**. 1. ed. Boca Raton: CRC, 2002. 738 p.
- SEBER, G. A. F.; WILD, C. J. **Nonlinear regression**. Hoboken: Wiley, 2003. 752 p.
- SERAFIM, M. et al. Sistema conservacionista e de manejo intensivo do solo no cultivo de cafeeiros na região do alto São Francisco, MG: um estudo de caso. **Bioscience Journal**, v. 27, n. 6, p. 964–977, Nov. 2011.
- SHEINER, L.; LUDDEN, T. Population Pharmacokinetics/pharmacodynamics. **Annual Review of Pharmacological Toxicology**, v. 8, p. 185–209, 1992.
- TEMESGEN, H.; MONLEON, V. J.; HANN, D. W. Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. **Canadian Journal of Forest Research**, Ottawa, v. 38, n. 3, p. 553–565, Mar. 2008.
- TORMENA, C. et al. Quantification of the soil physical quality of a tropical oxisol using the S index. **Scientia Agricola**, Piracicaba, v. 65, n. 1, p. 56–60, Feb. 2008.

VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S**. 4ed. ed. New York: Springer, 2002. 495 p.

WANG, Y.; LEMAY, V. M.; BAKER, T. G. Modelling and prediction of dominant height and site index of Eucalyptus globulus plantations using a nonlinear mixed-effects model approach. **Canadian Journal of Forest Research**, Ottawa, v. 37, n. 8, p. 1390–1403, Aug. 2007.

WILKINSON, G. N.; ROGERS, C. E. Symbolic description of factorial models for analysis of variance. **Journal of the Royal Statistical Society**. Series C (Applied Statistics), London, v. 22, n. 3, p. 392–399, Jan. 1973.

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

ZHAO, D.; WILSON, M.; BORDERS, B. E. Modeling response curves and testing treatment effects in repeated measures experiments: a multilevel nonlinear mixed-effects model approach. **Canadian Journal of Forest Research**, Ottawa, v. 35, n. 1, p. 122–132, Jan. 2005.

ANEXOS

ANEXO A Código R reproduzível correspondente ao ajuste do modelo van Genuchten reparametrizado. Disponível online em: <http://www.leg.ufpr.br/~walmes/TESE/anexoCRA.R>

```
##=====
## Uma parametrização do modelo van Genuchten para inferência
## sobre os parâmetros I e S.
##
##                               Walmes Marques Zeviani
##                               walmes@ufpr.br
##=====

##-----
## Definições da sessão.

require(lattice)
require(latticeExtra)
require(nlme)
require(car)
require(plyr)
require(reshape)

##-----
## Carrega os dados.

url <- "http://www.leg.ufpr.br/~walmes/data/ZevianiTeseCRA.txt"
cra <- read.table(url, header=TRUE, sep="\t")
cra$ltens <- log(cra$tens)
cra$ue <- with(cra, interaction(rep, PF, PS))
str(cra)

##-----
## Visualiza os dados.

xyplot(umid~ltens|PF*PS, groups=rep, data=cra)

##-----
## Define funções que retornam F e H para obter medidas de
## curvatura.

## modelo van Genuchten original, parametrização de Omuto 2007
vg.0 <- deriv3(~Ur+(Us-Ur)/(1+(exp(a)*exp(psi))^n)^(1-1/n),
              c("Ur", "Us", "a", "n"),
              function(psi, Ur, Us, a, n){ NULL })
```

```

## modelo van Genuchten reparametrizado para S e I
vg.SI <- deriv3(~Ur-S*(1+1/(1-1/n))^(1-1/n)+1)/
  (n*(1+exp(n*(psi-I))/(1-1/n))^(1-1/n)),
  c("Ur","S","I","n"),
  function(psi, Ur, S, I, n){ NULL })

##-----
## Ajusta as parametrizações por unidade experimental.

cra.ue <- split(cra, f=cra$ue)
str(cra.ue)

a0 <- lapply(cra.ue, nls,
  formula=umid~vg.0(ltens, Ur, Us, a, n),
  start=list(Ur=0.3, Us=0.6, a=0, n=1.5))
lapply(a0, summary)

aSI <- lapply(cra.ue, nls,
  formula=umid~vg.SI(ltens, Ur, S, I, n),
  start=list(Ur=0.3, S=-0.12, I=1.14, n=1.5))
lapply(aSI, summary)

##-----
## Medidas de curvatura.

lapply(a0, MASS::rms.curv)
lapply(aSI, MASS::rms.curv)

##-----
## Inferência sobre S e I pelo método delta.

B <- coef(a0[[1]])
W <- vcov(a0[[1]])
gS <- "-(Us-Ur)*n*(1+1/(1-1/n))^(1-1/n)-1"
gI <- "-a-log(1-1/n)/n"
deltaMethod(object=B, g=gS, vcov=W)
deltaMethod(object=B, g=gI, vcov=W)

dmSI <- function(nlsobj){
  B <- coef(nlsobj); W <- vcov(nlsobj)
  dmS <- unlist(deltaMethod(object=B, g=gS, vcov=W))
  dmI <- unlist(deltaMethod(object=B, g=gI, vcov=W))
  dmS <- c(dmS, lwr=dmS[1]-1.96*dmS[2], upr=dmS[1]+1.96*dmS[2])
  dmI <- c(dmI, lwr=dmI[1]-1.96*dmI[2], upr=dmI[1]+1.96*dmI[2])
  names(dmS) <- names(dmI) <- gsub("\\.\\.", "", names(dmS))
  return(data.frame(par=c("S","I"), rbind(dmS, dmI)))
}

vartheta <- lapply(a0, dmSI)
vartheta <- ldply(vartheta)
str(vartheta)

```

```

##-----
## Cria objeto de dados agrupados/longitudinais.

crag <- groupedData(umid~ltens|ue, data=cra)

##-----
## Ajuste do modelo não linear de efeitos mistos.
## van Genuchten original.

## Valores iniciais
start <- c(c(0.3, rep(0,9)), # Ur
           c(0.6, rep(0,9)), # Us
           c(0, rep(0,9)), # a
           c(1.8, rep(0,9))) # n

## Completo
nlme0.0 <- nlme(umid~Ur+(Us-Ur)/(1+exp(a+ltens)^n)^(1-1/n),
              data=crag,
              fixed=Ur+Us+a+n~PF*PS,
              random=Ur+Us~1,
              start=start)

## Efeitos aleatórios independentes
nlme0.1 <- update(nlme0.0, random=pdDiag(Ur+Us~1))

## Efeito aditivo para o parâmetro n
nlme0.2 <- update(nlme0.0, fixed=list(Ur+Us+a~PF*PS, n~PF+PS),
                 start=start[1:36])

anova(nlme0.0, nlme0.1) # testa correlação nula
anova(nlme0.0, nlme0.2) # testa interação nula em n

anova(nlme0.0, type="marginal") # teste de Wald no ef fixos
VarCorr(nlme0.0) # estivativas das variâncias

##-----
## Ajuste do modelo não linear de efeitos mistos.
## van Genuchten reparametrizado para S e I.

## Valores iniciais
start <- c(c(0.3, rep(0,9)), # Ur
           c(-0.1, rep(0,9)), # S
           c(1.6, rep(0,9)), # I
           c(1.8, rep(0,9))) # n

## Completo
nlmeSI.0 <- nlme(umid~Ur-S*(1+1/(1-1/n))^((1-1/n)+1)/
                (n*(1+exp(n*(ltens-I))/(1-1/n))^(1-1/n)),
                data=crag,
                fixed=Ur+S+I+n~PF*PS,
                random=Ur+S~1,
                start=start)

```

```
## Efeitos aleatórios independentes
nlmeSI.1 <- update(nlmeSI.0, random=pdDiag(Ur+S-1))

## Efeito aditivo para o parâmetro n
nlmeSI.2 <- update(nlmeSI.0, fixed=list(Ur+S+I~PF*PS, n~PF+PS),
                  start=start[1:36])

anova(nlmeSI.0, nlmeSI.1) # testa correlação nula
anova(nlmeSI.0, nlmeSI.2) # testa interação nula em n

anova(nlmeSI.0, type="marginal") # teste de Wald no ef fixos
VarCorr(nlmeSI.0)                # estimativas das variâncias

##-----
## Compara as verossimilhanças.

logLik(nlme0.0)
logLik(nlmeSI.0)

##-----
## Gráfico dos valores preditos.

plot(augPred(nlme0.0, level=0:1))
plot(augPred(nlmeSI.0, level=0:1))
plot(comparePred(nlme0.0, nlmeSI.0))

##-----
## sessionInfo()

## R version 2.15.3 (2013-03-01)
## Platform: i686-pc-linux-gnu (32-bit)

## [1] reshape_0.8.4      plyr_1.7.1           car_2.0-15
## [4] nnet_7.3-6          MASS_7.3-23         nlme_3.1-108
## [7] latticeExtra_0.6-24 RColorBrewer_1.0-5  lattice_0.20-13

##-----
```

ANEXO B Conteúdo de água do solo ($m^3 m^{-3}$) em função da tensão matricial (kPa), da profundidade (cm) e da unidade experimental (UE) para amostras coletadas na entre linha.

Prof. (cm)	UE	Tensão matricial (kPa)											
		0,01	1	2	4	6	10	33	100	500	1500		
0-20	1	0,57	0,57	0,56	0,52	0,49	0,45	0,40	0,36	0,35	0,34		
0-20	2	0,59	0,59	0,56	0,49	0,44	0,41	0,36	0,33	0,30	0,29		
0-20	3	0,59	0,59	0,57	0,50	0,46	0,42	0,37	0,35	0,35	0,34		
20-40	1	0,63	0,63	0,56	0,47	0,43	0,40	0,35	0,32	0,31	0,31		
20-40	2	0,61	0,59	0,55	0,43	0,39	0,36	0,32	0,29	0,28	0,27		
20-40	3	0,61	0,60	0,56	0,46	0,40	0,37	0,31	0,30	0,29	0,28		
40-80	1	0,57	0,57	0,55	0,47	0,42	0,39	0,35	0,32	0,30	0,30		
40-80	2	0,60	0,60	0,53	0,42	0,37	0,33	0,30	0,27	0,26	0,26		
40-80	3	0,58	0,58	0,52	0,42	0,38	0,35	0,31	0,29	0,27	0,27		
80-120	1	0,60	0,59	0,52	0,43	0,39	0,36	0,31	0,30	0,28	0,28		
80-120	2	0,61	0,58	0,52	0,42	0,39	0,35	0,31	0,29	0,28	0,27		
80-120	3	0,59	0,58	0,51	0,42	0,38	0,35	0,31	0,30	0,28	0,28		
120-160	1	0,62	0,57	0,50	0,44	0,41	0,38	0,33	0,31	0,28	0,27		
120-160	2	0,62	0,61	0,56	0,47	0,43	0,40	0,34	0,32	0,31	0,31		
120-160	3	0,55	0,54	0,51	0,47	0,44	0,42	0,36	0,35	0,34	0,33		

Disponível online em <http://www.leg.ufpr.br/~walmes/data/ZevianiTeseCRA.txt>.

ANEXO C Conteúdo de água do solo ($m^3 m^{-3}$) em função da tensão matricial (kPa), da profundidade (cm) e da unidade experimental (UE) para amostras coletadas na linha.

Prof. (cm)	UE	Tensão matricial (kPa)										
		0,01	1	2	4	6	10	33	100	500	1500	
0-20	1	0,64	0,61	0,57	0,44	0,39	0,36	0,31	0,28	0,26	0,25	
0-20	2	0,63	0,61	0,54	0,42	0,38	0,35	0,30	0,28	0,26	0,25	
0-20	3	0,66	0,65	0,57	0,44	0,39	0,35	0,30	0,28	0,26	0,27	
20-40	1	0,67	0,66	0,57	0,44	0,39	0,35	0,30	0,26	0,26	0,26	
20-40	2	0,64	0,60	0,54	0,40	0,36	0,33	0,30	0,27	0,24	0,24	
20-40	3	0,64	0,60	0,55	0,41	0,37	0,33	0,30	0,27	0,24	0,23	
40-80	1	0,62	0,61	0,51	0,39	0,35	0,31	0,27	0,26	0,24	0,24	
40-80	2	0,56	0,56	0,48	0,37	0,33	0,30	0,27	0,25	0,24	0,23	
40-80	3	0,61	0,60	0,51	0,38	0,34	0,30	0,28	0,25	0,24	0,24	
80-120	1	0,64	0,60	0,51	0,40	0,36	0,33	0,29	0,28	0,25	0,25	
80-120	2	0,61	0,56	0,49	0,41	0,38	0,34	0,31	0,29	0,27	0,27	
80-120	3	0,60	0,62	0,54	0,43	0,38	0,34	0,31	0,28	0,27	0,26	
120-160	1	0,61	0,59	0,54	0,44	0,38	0,35	0,30	0,29	0,28	0,28	
120-160	2	0,57	0,56	0,53	0,44	0,40	0,37	0,33	0,31	0,31	0,30	
120-160	3	0,57	0,54	0,49	0,41	0,38	0,36	0,33	0,31	0,29	0,28	

Disponível online em <http://www.ufpr.br/~walmes/data/ZevianiTeseCRA.txt>.