

Análise de modelos lineares mistos com dois fatores longitudinais: um quantitativo e um qualitativo ordinal

Marina Rodrigues Maestre ^{1 2}

César Gonçalves de Lima ³

Elias Silva de Medeiros ^{4 2}

Alinne da Silva ^{5 6}

Raffaella Rossetto ⁷

Resumo: Com o objetivo de encontrar um modelo que descreva o comportamento da variável *Capacidade de Troca de Cátions (CTC)* utilizou-se a classe dos modelos mistos. Consideraram-se dois fatores longitudinais: Tempo (quantitativo) e Profundidade (qualitativo ordinal). No modelo selecionado foram incluídos efeitos aleatórios no intercepto, no coeficiente linear e no quadrático em relação ao tempo, conseguindo explicar bem as correlações e a heterogeneidade de variâncias em diferentes níveis de profundidade e tempo.

Palavras-chave: *Profundidade, Tempo, Vinhaça, Variâncias diferentes.*

1 Introdução

Nos ensaios longitudinais é comum a realização de medidas repetidas considerando dois ou mais fatores longitudinais, como o tempo, profundidade, distância da linha de plantio etc.

Nesses ensaios também é comum admitir variâncias heterogêneas das variáveis respostas nos diversos níveis desses fatores longitudinais, além de correlações não nulas entre esses níveis. Esses aspectos devem ser levados em conta na formulação do modelo, pois podem alterar as inferências feitas sobre os parâmetros de interesse. Neste contexto, os modelos (lineares ou não lineares) mistos são indicados, pois favorecem a especificação de inúmeras estruturas de covariâncias que podem explicar bem a presença da correlação entre as medidas repetidas e a presença de heterogeneidade de variâncias.

O objetivo do presente trabalho foi utilizar a classe de modelos lineares mistos para buscar um modelo que se ajuste bem aos dados de um ensaio em blocos casualizados envolvendo dois fatores longitudinais: um fator quantitativo e um qualitativo ordinal.

¹ESALQ/USP. E-mail: ninamaestre@yahoo.com.br

²Agradecimento à CAPES pelo apoio financeiro.

³FZEA/USP. E-mail: cegdlima@usp.br

⁴ESALQ/USP. E-mail: eliasestatistica@gmail.com

⁵CENA/USP. E-mail: alinneagro@hotmail.com

⁶Agradecimento ao CNPq pelo apoio financeiro.

⁷APTA-SP. E-mail: rrossetto@apta.org.br

2 Material e métodos

2.1 Material

No presente estudo foram utilizados os dados de Capacidade de Troca de Cátions (CTC) de um experimento planejado num delineamento aleatorizado em blocos, com quatro repetições (SILVA, 2012). Os tratamentos foram: Vinhaça Não Concentrada (VNC) em três doses, 75, 150 e 300 m³ ha⁻¹, Vinhaça Concentrada (VC) em outras três doses, 7,5, 15 e 30 m³ ha⁻¹ e um tratamento controle, totalizando 7 tratamentos. As medidas de CTC foram feitas em três profundidades (0 a 0,25, 0,25 a 0,50 e 0,50 a 0,80 m) e em três tempos (2, 4 e 5 meses).

Os perfis individuais de respostas (Figura 1) evidenciam uma tendência quadrática nas medidas de CTC ao longo do tempo e do aumento da profundidade.

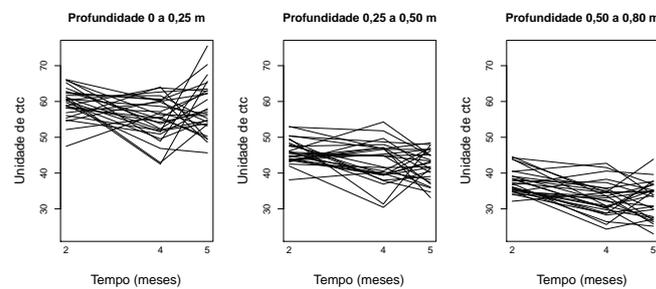


Figura 1: Perfis das respostas de CTC

2.2 Métodos

A classe dos modelos mistos admite estruturas de covariâncias diferentes para cada unidade experimental. O modelo linear misto utilizado para explicar o comportamento desses dados pode ser escrito como:

$$y_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \boldsymbol{\varepsilon}_{ij} \quad (1)$$

em que,

- \mathbf{y}_{ij} (9×1) é o perfil de respostas da parcela do i -ésimo bloco que recebeu o j -ésimo tratamento, sendo $i = 1, \dots, 4$ e $j = 1, \dots, 7$;
- \mathbf{X}_{ij} ($9 \times p$) é uma matriz de posto $p < 9$, conhecida e de especificação, associada ao vetor $\boldsymbol{\beta}$ ($p \times 1$) de parâmetros subpopulacionais desconhecidos;
- \mathbf{Z}_{ij} ($9 \times q$) é uma matriz de especificação conhecida, de posto coluna completo, associada ao vetor de efeitos aleatórios \mathbf{b}_{ij} ($q \times 1$) de diferenças individuais em torno dos valores populacionais, e
- $\boldsymbol{\varepsilon}_{ij}$ (9×1) é um vetor de erros aleatórios associado a \mathbf{y}_{ij} .

As matrizes de especificação \mathbf{X}_{ij} e \mathbf{Z}_{ij} podem ser diferentes e variar entre unidades experimentais. Supõe-se também que \mathbf{b}_{ij} e $\boldsymbol{\varepsilon}_{ij}$ são independentes, sendo $\boldsymbol{\varepsilon}_{ij} \sim N_9(\mathbf{0}, \mathbf{R}_{ij})$ e $\mathbf{b}_{ij} \sim N_q(\mathbf{0}, \mathbf{G})$, em que, \mathbf{R}_{ij} (9×9) e \mathbf{G} ($q \times q$) são matrizes de covariâncias.

O modelo (1) pode ser formulado em dois estágios, em que são identificadas as características individuais e populacionais. No primeiro estágio, para cada unidade experimental (ij), tem-se:

$$\mathbf{y}_{ij} | \mathbf{b}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \boldsymbol{\varepsilon}_{ij} \sim N(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{b}_{ij}, \mathbf{R}_{ij}), \quad (2)$$

em que, \mathbf{R}_{ij} é conhecida como matriz de dispersão condicional e está associada ao erro condicional $\boldsymbol{\varepsilon}_{ij} = \mathbf{y}_{ij} - \mathbf{X}_{ij}\boldsymbol{\beta} - \mathbf{Z}_{ij}\mathbf{b}_{ij}$.

No segundo estágio, assume-se que $\mathbf{b}_{ij} \sim N(\mathbf{0}, \mathbf{G})$ é independente de $\boldsymbol{\varepsilon}_{ij}$, obtendo-se o modelo marginal (ou não condicional):

$$\mathbf{y}_{ij} \sim N(\mathbf{X}_{ij}\boldsymbol{\beta}, \mathbf{Z}_{ij}\mathbf{G}\mathbf{Z}_{ij}^t + \mathbf{R}_{ij}) \quad (3)$$

em que, a matriz $\mathbf{V}_{ij} = \mathbf{Z}_{ij}\mathbf{G}\mathbf{Z}_{ij}^t + \mathbf{R}_{ij}$ é chamada matriz de dispersão marginal e está associada ao erro marginal $\mathbf{e}_{ij} = \mathbf{y}_{ij} - \mathbf{X}_{ij}\boldsymbol{\beta}$.

Segundo Suyama (1995), o modelo (3) envolve três modelos: um para os efeitos fixos ($\mathbf{X}_{ij}\boldsymbol{\beta}$) que deve descrever a(s) curva(s) média(s) de resposta; um para os efeitos aleatórios ($\mathbf{Z}_{ij}\mathbf{b}_{ij}$) que, juntamente com os efeitos fixos, devem descrever as curvas individuais de respostas; e um para a matriz de dispersão condicional (\mathbf{R}_{ij}) que deve descrever as covariâncias intra-indivíduos ao longo das profundidades e dos tempos de observação.

No modelo marginal, os parâmetros desconhecidos em \mathbf{G} e \mathbf{R}_{ij} compõem o vetor de parâmetros de variâncias e covariâncias, $\boldsymbol{\theta}$, e são estimados com base no método da máxima verossimilhança.

Na seleção do melhor modelo será utilizado o teste de Wald para avaliar a significância dos efeitos fixos. O teste da razão de verossimilhanças (TRV) será usado para selecionar efeitos fixos e aleatórios em modelos hierárquicos ou encaixados, por meio da comparação de modelos simples com modelos mais gerais.

Nobre e Singer (2007) sugerem a utilização do gráfico de quantis normais com envelope simulado para estudar a normalidade dos erros condicionais padronizados. A sua homocedasticidade é avaliada no gráfico desses erros *versus* os valores ajustados pelo modelo. A presença de observações discrepantes é avaliada num gráfico desses resíduos *versus* os índices das observações.

Para o ajuste dos modelos lineares mistos utilizou-se a função `lme` da biblioteca `nlme` (PINHEIRO; BATES, 2000) do programa R (R CORE TEAM (2013)).

3 Resultados e discussão

Os modelos ajustados para os dados da variável CTC estão na Tabela 1. O modelo *m3.ml* foi ajustado devido ao resultado dos efeitos significativos do modelo *m2.ml* pelo teste de Wald.

Tabela 1: Modelos ajustados e teste da Razão de Verossimilhanças

| Modelo | Efeito fixo (*) | Efeito aleatório | ℓV | er | TRV | Valor p |
|---------------------|-------------------------------------|---------------------------------------|----------|----------------|----------------|-----------|
| <i>m.r eml</i> | interações | intercepto, Tempo | 1249,103 | | | |
| <i>m.r eml</i> | interações | intercepto, Tempo, Tempo ² | 1264,754 | <i>m.r eml</i> | <i>m.r eml</i> | 0,0013 |
| <i>m.ml</i> | interações | intercepto, Tempo, Tempo ² | 1382,208 | | | |
| <i>m.ml</i> | Prof Tempo, ProfT empo ² | intercepto, Tempo, Tempo ² | 1440,541 | <i>m.ml</i> | <i>m.ml</i> | 0,1459 |
| <i>m.r eml</i> | Prof Tempo, ProfT empo ² | intercepto, Tempo, Tempo ² | 1422,411 | | | |
| <i>m.r eml</i> (**) | Prof Tempo, ProfT empo ² | intercepto, Tempo, Tempo ² | 1390,247 | <i>m.r eml</i> | <i>m.r eml</i> | <0,0001 |

(*) Intercepto, Bloco, Tratamento, Profundidade, Tempo, Tempo²

(**) Variâncias diferentes para cada Profundidade e para cada Tempo

Assim, pelos testes mostrados na Tabela 1, o modelo final é *m4.reml*:

$$ctc_{ijkl} = [\beta_0 + b_0] + \beta_1 Bloco_i + \beta_2 Trat_j + \beta_3 Prof_k + [\beta_4 + b_{4l}] Tempo_l + [\beta_5 + b_{5l}] Tempo_l^2 + \beta_6 (Prof \times Tempo)_{kl} + \beta_7 (Prof \times Tempo^2)_{kl} + \varepsilon_{ijkl}.$$

Os gráficos de diagnósticos de resíduos da Figura 2 mostram que o modelo está bem ajustado. O teste Shapiro-Wilk deu um Valor p de 0,953, indicando que esses se aproximam da distribuição normal.

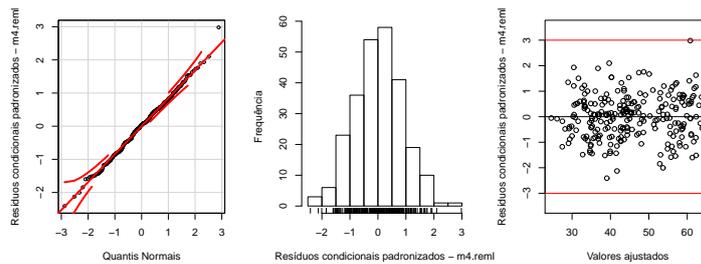


Figura 2: Diagnósticos dos resíduos do modelo *m4.reml*

As curvas médias e individuais ajustadas para cada tratamento e cada profundidade ao longo do tempo, bem como as observações da referida variável resposta estão nos gráficos da Figura 3. Pode-se observar que o modelo explicou bem o comportamento médio da variável CTC dos diferentes tratamentos em função do aumento das profundidades e do tempo de coleta.

4 Conclusão

Considerando o método de estimação *REML*, o melhor modelo ajustado aos dados observados foi:

- na parte fixa, com efeitos de bloco, tratamento, profundidade, tempo e tempo² e as interações profundidade \times tempo e profundidade \times tempo²,
- com efeitos aleatórios incluídos no Intercepto, tempo e tempo², e
- com diferentes variâncias para cada profundidade e diferentes variâncias para cada tempo.

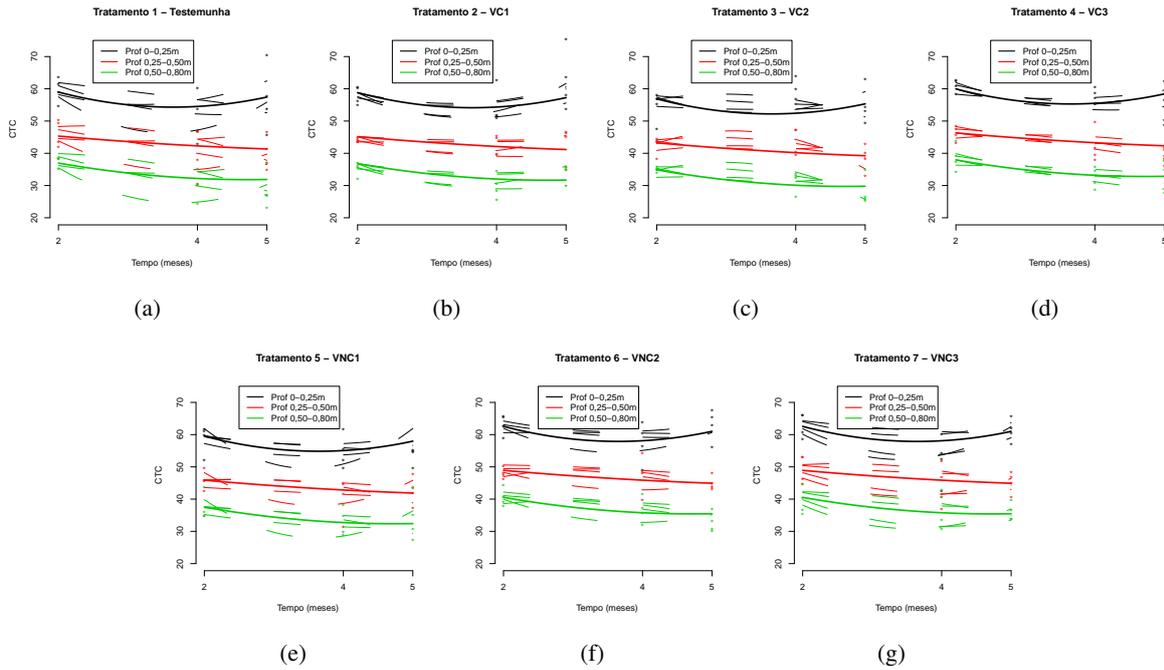


Figura 3: Curvas médias e individuais e observações da variável CTC para o modelo $m4.reml$

5 Algumas considerações

Uma vez ajustado o modelo que represente bem o comportamento da variável CTC, deve-se prosseguir a análise, comparando-se os efeitos dos tratamentos através da comparação dos parâmetros das curvas quadráticas.

Referências

- [1] NOBRE, J. S.; SINGER, J. M. Residual Analysis for Linear Mixed Models. *Biometrical Journal*, Weinheim, v. 49, n. 6, p. 863-875, Jan., 2007.
- [2] PINHEIRO, J. C.; BATES, D. M. *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag, 528p. 2000.
- [3] R CORE TEAM. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-13-5, URL <http://www.R-project.org/> 2013.
- [4] SILVA, A. *Vinhaça concentrada de cana-de-açúcar: monitoramento das propriedades químicas do solo e mineralização líquida de nitrogênio*. 2012. 107 p. Tese (Doutorado) - CENA-USP, Piracicaba, 2012.
- [5] SUYAMA, E. *Identificação de um modelo de efeitos aleatórios*. 1995. 129 p. Tese (Doutorado) - IME-USP, São Paulo, 1995.