

Programa R para seleção de preditores em modelos de superfície de resposta

Ricardo de Almeida Pazianotto¹

Aline de Holanda Nunes Maia¹

1 Introdução

Superfícies de resposta são modelos empíricos amplamente utilizados para descrever padrões de resposta de variáveis de interesse a variações em preditores contínuos. Um tipo de particular interesse para dados agrícolas é a superfície de resposta quadrática cujos preditores são os efeitos lineares dos fatores A e B, sua interação (AxB) e respectivos efeitos quadráticos (EDMONDSON, 1991; BARROS, 1999). Em experimentos com um grande número de variáveis-resposta, é comum que para algumas delas, a contribuição de algum efeito seja negligível, o que requer exclusão de preditores para obtenção de um modelo reduzido (PETTER et al, 2012). Os métodos tradicionais de seleção de preditores utilizados em regressão múltipla, como por exemplo, o *stepwise*, não podem ser empregados nesse caso, devido à natureza hierárquica dos efeitos em modelos de superfície de resposta: os termos lineares não podem ser excluídos sempre que a interação ou respectivos termos quadráticos permaneçam no modelo (MACCULLAGH & NELDER, 1983). Outro aspecto importante é alteração dos valores p associados aos parâmetros do modelo cada vez que ocorre a exclusão de um preditor, em decorrência da não ortogonalidade das colunas da matriz do delineamento. Neste trabalho, discutimos algumas limitações de procedimentos do SAS/STAT[®] e pacotes do R utilizados na análise de superfícies de resposta e apresentamos um programa R para seleção de preditores em modelos mistos onde a parte do modelo correspondente aos fatores fixos é representada por uma superfície de resposta. É apresentado exemplo de seleção de preditores num modelo de superfície de resposta quadrática, para dados provenientes de um delineamento casualizado em blocos. São discutidas limitações e possíveis extensões do algoritmo para modelos mistos com estruturas mais complexas associadas aos fatores aleatórios.

¹ Embrapa Meio Ambiente. E-mail: ricardo.pazianotto@embrapa.br

2 Material e métodos

Desenvolvemos um programa R para seleção de preditores em modelos de superfície de resposta em modelos mistos onde a parte fixa é representada por um modelo polinomial. O programa é constituído de duas partes: a) uma função desenvolvida para seleção de preditores (*step.rs*) a partir dos resultados de análise de variância da parte fixa do modelo completo e b) códigos que utilizam o *nlme package* para análise do modelo misto (**Figura 1**).

A seleção de preditores ocorre de acordo com o seguinte protocolo: a) os termos do modelo, exceto o intercepto, são ordenados em ordem decrescente de nível de significância nominal (valor p); b) O termo correspondente ao maior valor é excluído, se este for superior a um nível crítico estabelecido, exceto se este for um termo linear com termos de ordem superior ainda presentes no modelo; c) se o máximo valor p for inferior ao valor crítico, seleciona-se o modelo atual; d) caso contrário, o repete-se o procedimento.

Apresentamos um exemplo de um ensaio para avaliar o efeito de lâminas de água (W, mm) e adubação nitrogenada (N, kg/ha) na produtividade do meloeiro no Vale do Curu, Ceará. Os dados foram simulados a partir dos resultados apresentados em BARROS et al, 1999. Foi ajustado um modelo misto onde a parte fixa é a superfície de resposta quadrática com os fatores N e W e a parte aleatória o fator bloco.

3 Resultados e discussão

A função para seleção de preditores (*step.sr*), foi desenvolvida segundo o protocolo descrito no item Material e Métodos. Os códigos R para ajuste do modelo misto utilizando esta função são apresentados na Figura 1. Alguns pacotes do R e procedimentos do SAS/STAT têm funcionalidades similares, no entanto, apresentam ainda algumas limitações para o caso de superfícies de resposta polinomiais. No R, temos as funções *step* do pacote *stats* e *stepAIC* do pacote *MASS* que selecionam o modelo utilizando o método *stepwise* com opções *backward*, *forward* e *both*; no *REG Procedure* do SAS/STAT®, há a opção *selection* para definição do método de seleção de preditores: *stepwise*, *backward selection* ou *forward selection*. Esses procedimentos, no entanto, não se adequam a modelos de superfície de resposta em função das restrições de permanência dos termos lineares sempre que interações ou termos correspondentes de maior ordem estejam presentes no modelo. No *RSREG Procedure* do SAS/STAT, procedimento específico para ajuste de superfícies de resposta, não há opções para redução do modelo; o pacote correspondente do R, o “*rms package*”, possui funções especiais para inclusão de interações duplas e termos de ordem quadrática de forma conjunta, por exemplo, a inclusão é feita para todos os termos quadráticos. Em alguns

problemas, no entanto, a resposta pode ser quadrática para um dos fatores e linear para outro, o que inviabiliza seu uso.

O R possui alguns pacotes para seleção automatizada de modelos como, por exemplo, o pacote *MuMIn* que possui a função *dredge* que gera todos os modelos possíveis para uma combinação de preditores e classifica-os de acordo com algum critério de informação (AIC, BIC, etc). Esta função respeita as restrições de marginalidade, não incluindo modelos que contenham interação sem seus respectivos efeitos principais, além disso, outras restrições podem ser incluídas através das opções *subset* e *fixed*. Quando utilizamos esta função na seleção de modelos para superfície resposta, devemos incluir uma restrição para cada variável preditora a fim de excluir modelos que tenham o termo quadrático mas não o linear. Uma limitação, no entanto, é o tempo computacional, pois o número de combinações possíveis cresce exponencialmente com o número de preditores.

```
# Carregando pacotes
# pacote para análise de modelos mistos
require(nlme)
# pacote para leitura do arquivo de dados (planilha Excel)
require(xlsReadWrite)

# Definindo o diretório de trabalho
setwd("C:/dados")

# Importando o conjunto de dados
melao=read.xls('melao.xls')
melao

# Ajustando modelo (Exemplo: delineamento casualizado em blocos, fatores W e N)
fit=lme(prod ~ N*W + I(N^2)+ I(W^2), data = melao, random = ~ 1|B1, method="ML")
summary(fit1)

# Reduzindo o modelo com step.rs
(aj1=step.rs(fit))
summary(aj1)
```

Figura 1. Códigos R para análise do modelo misto e seleção de preditores (redução da parte fixa do modelo) utilizando a função *step.rs*.

Para o exemplo que aborda o efeito de lâminas de água (W) e adubação nitrogenada (N) sobre a produtividade do meloeiro, obtivemos um modelo reduzido com os termos lineares e quadráticos para N e W, sem interação, utilizando a função *step.rs* aqui apresentada. Quando

usamos a função *stepAIC*, o efeito linear de N é excluído, além da interação, pelo fato de esse procedimento não respeitar as restrições hierárquicas válidas para seleção de preditores em superfícies de resposta. As análises de variância da parte fixa do modelo são apresentadas na **Tabela 1**. Procedimentos como o *REG* e o *RSREG* do *SAS* e o *step* do pacote *MASS* do *R* não são adequados nesse caso, pois não permitem a análise de modelos mistos.

Tabela 1. Análise de variância da parte fixa dos modelos reduzidos a partir do modelo completo, quando são utilizadas as funções *stepAIC* e *step.rs*. O modelo completo é representado por uma superfície de resposta quadrática com os fatores lâmina de água (W, mm) e adubação nitrogenada (N, kg/ha).

Função	Preditor	Estimativa	Erro-padrão	Estatística t	Valor p
<i>stepAIC</i>	Intercepto	1,584	1,404	1,128	0,2618
	W	0,0617	0,009	6,895	<0,0001
	N ²	-6,3E-05	5,3E-06	-11,9823	<0,0001
	W ²	-4,5E-05	1,25E-05	-3,5888	0,0005
<i>step.rs</i>	Intercepto	1,368	1,490	0,918	0,3606
	N	0,00348	0,00781	0,445	0,6569
	W	0,061713	0,008981	6,871	<0,0001
	N ²	-7,1E-05	1,87E-05	-3,814	0,0002
	W ²	-4,5E-05	1,26E-05	-3,576	0,0005

Nesse exemplo, apenas o bloco foi tratado como fator aleatório, o que simplifica a redução do modelo fixo. Em casos onde a estrutura de fatores aleatórios é mais complexa ou há presença de medidas repetidas, fazendo com que a modelagem das estruturas de covariância seja parte importante do problema, será necessário, a cada passo de redução da parte fixa, considerar também a escolha de estruturas de covariância mais adequadas. Esse problema é abordado em FERNANDEZ, 2007.

4 Conclusões

O programa *R* desenvolvido é útil para seleção de preditores em modelos mistos cuja parte fixa é representada por superfícies de resposta, em situações onde a modelagem de estruturas de covariância associadas aos fatores aleatórios ou estrutura de medidas repetidas não é o interesse principal do estudo. Agiliza o processo de redução dos modelos, particularmente em ensaios com um grande número de variáveis-resposta de interesse, onde para alguns termos da parte fixa do modelo, não há evidência de contribuição relevante na modelagem do processo.

5 Bibliografia

- [1] BARROS, V. da S. **Função de produção do melão (*Cucumis melo* L.) aos níveis de água e adubação nitrogenada no Vale do Curu-CE.** 1999. 91p. Dissertação (Mestrado) – Universidade Federal do Ceará, Fortaleza, 2009.
- [2] EDMONDSON, R. N. Agricultural response surface experiments based on four-level factorial designs. **Biometrics**, v.47, p.1435–1448, 1991.
- [3] FERNANDEZ, G. Model Selection in PROC MIXED - A User-friendly SAS® Macro Application. Disponível em: <http://www2.sas.com/proceedings/forum2007/191-2007.pdf>
Acesso em: 10 de abril de 2013.
- [4] MADAMBA, P. S. The Response Surface Methodology: An Application to Optimize Dehydration Operations of Selected Agricultural Crops. **LWT - Food Science and Technology**, v. 35(7), p.584–592, 2002.
- [5] PETTER, F. A; B. E. MADARI; M. A. S. DA SILVA; M. A. C. CARNEIRO; M. T. DE M. CARVALHO; B. H. MARIMON JÚNIOR; L. P. PACHECO. Soil fertility and upland rice yield after biochar application in the Cerrado. **Pesq. Agropec. Bras.**, Brasília, v. 47, n. 5, 2012 .