Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Análise Bayesiana de ensaios fatoriais 2^k usando os princípios dos
efeitos esparsos, da hierarquia e da hereditariedade

Guilherme Biz

Dissertação apresentada para obtenção do título de Mestre
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba
2009

Guilherme Biz
Bacharel em Matemática

**Análise Bayesiana de ensaios fatoriais 2^k usando os princípios dos efeitos
esparsos, da hierarquia e da hereditariedade**

Orientador:
Prof. Dr. **SILVIO SANDOVAL ZOCCHI**

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Es-
tatística e Experimentação Agronômica

Piracicaba
2009

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Biz, Guilherme

Análise Bayesiana de ensaios fatoriais 2^k usando os princípios dos efeitos esparsos, da hierarquia e da hereditariedade / Guilherme Biz. - Piracicaba, 2009.
71 p. : il.

Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", 2009.
Bibliografia.

1. Análise de regressão e de correlação 2. Estatística aplicada 3. Inferência Bayesiana
Título

CDD 519.542
B625a

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

Dedicatória

Aos meus pais, Valdir e Sandra,
à minha irmã, Milena,
à minha namorada, Claudia,
e ao meu avô, Pascoal (in memoriam).

AGRADECIMENTOS

Primeiramente a Deus, que me deu força para concluir esta dissertação.

Agradeço o meu orientador Silvio Sandoval Zocchi por ter me ajudado muito, mesmo em momentos difíceis, mas finalizamos o trabalho. Aos professores do departamento, em especial a professora Roseli Aparecida Leandro, pela grande ajuda, realizando o meu desejo de concretizar este trabalho.

Agradeço a todos os meus amigos desta pós graduação pela ajuda e os bons momentos juntos : Bruno, Cássio, Eduardo, Elisabete, Elton, Epaminondas, Evandro, Fernanda, Juliana, Júlio, Lucimary, Luiz Alberto, Luiz Trench, Marcelino, Mariana, Michele, Paula Fonte, Raphael, Renata, Renato, Rodrigo, Simone, Vanderly e Wilson

Agradeço as secretárias Solange, Luciane e Eduardo que sempre me ajudaram tirando dúvidas e resolvendo as pendências burocráticas.

Agradeço muito a minha família, que me ajudou muito em momentos difíceis e importantes para eu realizar este trabalho: Meus pais, Valdir e Sandra; minha irmã Milena.

Agradeço em especial a minha namorada Claudia, por me aguentar longe e as vezes nervoso.

E a todos aqueles que acreditaram em mim.

OBRIGADO!!!

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 REVISÃO DE LITERATURA	19
3 MATERIAIS E MÉTODOS	25
3.1 Materiais	25
3.2 Métodos	27
4 RESULTADOS E DISCUSSÃO	39
4.1 Dados Simulados	39
4.2 Dados de biofilme	48
5 CONCLUSÕES	61
REFERÊNCIAS	63
ANEXOS	65

RESUMO

Análise Bayesiana de ensaios fatoriais 2^k usando os princípios dos efeitos esparsos, da hierarquia e da hereditariedade

No planejamento de experimentos para o ajuste de modelos polinomiais envolvendo k fatores principais e respectivas interações, é bastante comum a utilização dos fatoriais 2^k , 3^k ou frações dos mesmos. Para as análises dos resultados desses experimentos, frequentemente se considera o princípio da hereditariedade, ou seja, uma vez constatada uma interação significativa entre fatores, os fatores que aparecem nesta interação e respectivas interações devem também estar presentes no modelo. Neste trabalho, esse princípio é incorporado diretamente à priori, para um método de seleção de variáveis Bayesiana, seguindo as idéias propostas por Chipman, Hamada e Wu (1997), porém com uma alteração dos valores sugeridos pelos autores para os hiperparâmetros. Essa alteração, proposta neste trabalho, promove uma melhoria considerável na metodologia original. A metodologia é então ilustrada por meio da análise dos resultados de um experimento fatorial para a elaboração de biofilmes de amido originado da ervilha.

Palavras-chave: Análise Bayesiana; Princípio da hereditariedade; Princípio dos efeitos esparsos; princípio da hierarquia; Ensaio fatorial

ABSTRACT

Bayesian analysis of 2^k factorial designs using the sparse effects, hierarchy and heredity principles

In experimental planning for adjustment of polynomials models involving k main factors and their interactions, it is frequent to adopt the 2^k , 3^k designs or its fractions. Furthermore, it is not unusual, when analysing the results of such experiments, to consider the heredity principle. In other words, once detected a significant interaction between factors, the factors that appear in this interaction and respective interactions should also be present in the model. In this work, this principle is incorporated directly in the prior, following the ideas proposed by Chipman, Hamada and Wu (1997), but changing some of the hyperparameters. What improves considerably the original methodology. Finally the methodology is illustrated by the analysis of the results of an experiment for the elaboration of pea starch biofilms.

Keywords: Bayesian analysis; Heredity principle; Sparse effects principle; Hierarchy principle; Factorial design

LISTA DE FIGURAS

Figura 1 - Esquema de dependência do termo de quarta ordem A^2B^2 segundo o princípio da hereditariedade	29
Figura 2 - Esquema de dependência dos termos de modelos polinomiais contendo três fatores principais e interações de ordem até três, segundo o princípio da hereditariedade	30
Figura 3 - Probabilidades a priori dos 14 modelos mais prováveis de acordo com a intensidade da hereditariedade	37
Figura 4 - Índice H de diversidade dos modelos a posteriori segundo o valor do hiperparâmetro τ e da intensidade da hereditariedade: fraca, fraca rigorosa, forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*)	40
Figura 5 - Probabilidades a posteriori de alguns modelos mais prováveis considerando-se diferentes valores de τ , hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*) , em que os números de 1 a 9 correspondem aos modelos descritos na tabela 7	43
Figura 6 - Probabilidades a posteriori dos termos considerados no modelo serem ativos, considerando-se diferentes valores de τ e hereditariedade fraca, fraca rigorosa (fr.rig.) e forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*)	45
Figura 7 - Densidades a posteriori dos parâmetros, para análise do dados simulados, considerando-se as hereditariedades fraca, fraca rigorosa e forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho e o intervalo de 90% de credibilidade (HPD)	47
Figura 8 - Cadeias de parâmetros, para a análise dos dados simulados, considerando-se as hereditariedades fraca, fraca rigorosa e forte proposta por Chipman, Hamada e Wu(1997) e forte proposta neste trabalho	49

Figura 9 - Índice H de diversidade dos modelos a posteriori, para os dados de biofilme, segundo o valor do hiperparâmetro τ e a intensidade da hereditariedade: fraca, fraca rigorosa ou forte proposta neste trabalho (forte*)	51
Figura 10 - Probabilidades a posteriori de alguns modelos mais prováveis, para os dados de biofilme, considerando-se diferentes valores de τ , hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*), em que os números de 1 a 7 representam os modelos descritos na tabela 12	53
Figura 11 - Probabilidades a posteriori dos termos considerados no modelo serem ativos, para os dados de biofilme, considerando-se diferentes valores de τ e hereditariedades fraca, fraca rigorosa (fr.rig.) e forte proposta neste trabalho (forte*), descritos na tabela 13	56
Figura 12 - Cadeias de parâmetros para a análise dos dados de biofilme, considerando-se as hereditariedades fraca, fraca rigorosa e forte proposta neste trabalho . . .	57
Figura 13 - Densidades a posteriori dos parâmetros, para a análise dos dados de Biofilme, para as hereditariedades fraca, fraca rigorosa e forte proposta neste trabalho e o intervalo de 90% de credibilidade (HPD)	59

LISTA DE TABELAS

Tabela 1 - Níveis dos fatores A , B e C de um delineamento fatorial completo 2^3 e coeficientes dos contrastes ($- = -1$ e $+ = +1$) para a estimação dos efeitos de interação dupla e tripla	20
Tabela 2 - Níveis dos fatores A e B de um delineamento fatorial completo 3^2 , e coeficientes dos contrastes para a estimação dos efeitos de primeira e segunda ordem em modelos polinomiais de segunda ordem completo	20
Tabela 3 - Níveis dos fatores do delineamento 2^3 completo, e observações da variável resposta y , obtidas por meio de simulação	25
Tabela 4 - Formulações de diferentes biofilmes de ervilha, coeficientes para estimação das interações e alongamentos sob força máxima (y) dos biofilmes formados	26
Tabela 5 - Sugestões de valores para os hiperparâmetros p_{00} , p_{01} , p_{10} e p_{11} da distribuição a priori de δ propostos por Chipman, Hamada e Wu (1997) e novos valores propostos neste trabalho (*) para uma hereditariedade forte	35
Tabela 6 - Probabilidades a priori de cada coeficiente do modelo 30 ser importante de acordo com a intensidade da hereditariedade, cujos hiperparâmetros são apresentados na Tabela 5; Números k de possíveis modelos e índice de diversidade H dos possíveis modelos	36
Tabela 7 - Probabilidades a posteriori de alguns modelos mais prováveis considerando-se diferentes valores de τ , hereditariedades fraca, fraca rigorosa (fr. rig.) e forte propostas por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*), números de diferentes modelos (k) e índices de diversidade de Shannon (H). Em negrito são ressaltados os modelos mais prováveis para cada valor de τ e cada intensidade de hereditariedade	42
Tabela 8 - Probabilidades a posteriori dos termos considerados no modelo serem ativos, considerando-se diferentes valores de τ e hereditariedades fraca, fraca rigorosa (fr.rig.) e forte propostas por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*)	44

Tabela 9 - Estimativas dos parâmetros do modelo considerando-se $\tau = \tau^*$ e hereditariedades fraca, fraca rigorosa (fr. rig.) e forte propostas por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*)	46
Tabela 10 -Estatística (Z) para o teste de convergência de Geweke (1992) considerando-se $\tau = \tau^*$ e a hereditariedade fraca, fraca rigorosa (fr. rig.) e forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*),	48
Tabela 11 -Estatísticas (FD) para o teste de convergência de Raftery e Lewis (1992) considerando-se $\tau = \tau^*$ e hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*)	50
Tabela 12 -Probabilidades a posteriori de alguns modelos mais prováveis, para o experimento do Biofilme, considerando-se diferentes valores de τ , hereditariedade fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*), números de diferentes modelos (k) e índices de diversidade de Shannon (H), ressaltando-se, em negrito, o modelo mais provável em cada uma das situações	52
Tabela 13 -Probabilidades a posteriori dos termos considerados no modelo serem ativos, para os dados de biofilme, considerando-se diferentes valores de τ e hereditariedades fraca, fraca rigorosa (fr.rig.) e forte proposta neste trabalho (forte*)	55
Tabela 14 -Estatística (Z) para o teste de convergência de Geweke (1992), para o experimento do biofilme, considerando-se $\tau = \tau^*$ e a hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*)	58
Tabela 15 -Estatística (FD) para o teste de convergência de Raftery e Lewis (1992), para os dados de biofilme, considerando-se $\tau = \tau^*$ e as hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*)	58
Tabela 16 -Estimativas dos parâmetros do modelo, para os dados de biofilme, considerando-se $\tau = \tau^*$ e as hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*)	60

1 INTRODUÇÃO

Em diversas áreas de pesquisa, frequentemente o pesquisador deseja obter uma função matemática que relaciona uma variável com uma ou mais variáveis ou fatores de interesse. Especificamente, ele busca determinar um subconjunto de variáveis explicativas para melhor explicar a variável resposta. Quando existem muitas variáveis explicativas e consequentemente muitos parâmetros no modelo, geralmente são usados métodos automáticos para a seleção de variáveis. Esses métodos, entretanto, não contemplam princípios comumente adotados nesses casos, como o dos efeitos esparsos, o da hierarquia e o da hereditariedade. Este último, considera que, uma vez constatada uma interação entre fatores significativa, considera-se que os fatores que aparecem nessa interação e suas interações devem também estar presentes no modelo. Este trabalho utiliza a aproximação Bayesiana considerando o princípio de hereditariedade, incorporado diretamente na escolha da distribuição a priori seguindo as idéias propostas por Chipman, Hamada e Wu (1997).

A metodologia é, então, ilustrada por meio da análise dos resultados de um experimento para a elaboração de biofilme de amido originado da ervilha e de um conjunto de dados obtidos por meio de simulação.

2 REVISÃO DE LITERATURA

Em pesquisas científicas frequentemente há o interesse no estudo da influência de dois ou mais fatores simultaneamente sobre uma ou mais variáveis resposta de interesse. Para isso, geralmente são utilizados os delineamentos fatoriais, que consideram todos os níveis de cada fator e todas suas possíveis combinações como tratamentos. Se forem considerados k fatores, cada um com N níveis, tem-se o esquema fatorial completo denotado por N^k , com N^k tratamentos.

Dois casos especiais são particularmente importantes e frequentes, os que ocorrem quando os k fatores tem apenas dois ($N = 2$) ou três ($N = 3$) níveis, denotados por 2^k e 3^k , respectivamente.

Os fatoriais 2^k são especialmente usados para a seleção de fatores importantes sobre a variável, dentre os k fatores candidatos. Apresentam a grande vantagem de possuírem um número menor de observações do que outros delineamentos com k fatores. Devido ao fato de os fatores possuírem apenas dois níveis, são usados para estudar o efeito linear dos níveis dos fatores sobre a resposta, e segundo Myers e Montgomery (1995), são utilizados, ainda, como parte de outros delineamentos, como o delineamento composto central. A segunda, terceira e quarta colunas da tabela 1 apresentam, como ilustração, os níveis dos fatores de um fatorial completo 2^3 , em que $-$ representa um nível e $+$, o outro nível. Esses sinais são usados também, na estimação dos efeitos dos fatores e respectivas interações sobre a variável resposta, considerando-os como os sinais dos coeficientes dos contrastes envolvendo as respostas. Os sinais para o cálculo dos efeitos das interações, por sua vez, são calculados por meio da multiplicação dos sinais de seus efeitos formadores.

Em situações em que há o interesse de se ajustar modelos polinomiais de segunda ordem, por outro lado, geralmente são utilizados delineamentos considerando fatores com mais de dois níveis, por exemplo, três níveis. Com isso, são apresentados os fatoriais da série 3^k , em que cada fator possui 3 níveis. Um exemplo dos fatoriais 3^k é o delineamento 3^3 que é muito usado em pesquisas em ensaios de adubação N-P-K, que possui 3 fatores: nitrogênio (N), fósforo (P) e potássio (K), cada um com 3 níveis $(-1, 0, 1)$. Neste caso -1 representa a dose mais baixa do nutriente, 0, a intermediária e $+1$, a mais alta, totalizando $3^3 = 27$ tratamentos. No caso de $k = 2$ nutrientes com três níveis cada, tem-se o delineamento 3^2 com

Tabela 1 - Níveis dos fatores A , B e C de um delineamento fatorial completo 2^3 e coeficientes dos contrastes ($- = -1$ e $+ = +1$) para a estimação dos efeitos de interação dupla e tripla

Tratamentos	A	B	C	AB	AC	BC	ABC
1	-	-	-	+	+	+	-
2	-	-	+	+	-	-	+
3	-	+	-	-	+	-	+
4	-	+	+	-	-	+	-
5	+	-	-	-	-	+	+
6	+	-	+	-	+	-	-
7	+	+	-	+	-	-	-
8	+	+	+	+	+	+	+

9 tratamentos, apresentado na tabela 2.

Tabela 2 - Níveis dos fatores A e B de um delineamento fatorial completo 3^2 , e coeficientes dos contrastes para a estimação dos efeitos de primeira e segunda ordem em modelos polinomiais de segunda ordem completo

Tratamentos	A	B	AB	A^2	B^2
1	-1	-1	1	1	1
2	-1	0	0	1	0
3	-1	1	-1	1	1
4	0	-1	0	0	1
5	0	0	0	0	0
6	0	1	0	0	1
7	1	-1	-1	1	1
8	1	0	0	1	0
9	1	1	1	1	1

A codificação apresentada nas tabelas 1 e 2 são feitas para facilitar, tanto a

construção do delineamento experimental, quanto a análise posterior dos resultados do experimento. Considere um fator quantitativo cujos níveis pertencem a um certo intervalo. Uma codificação usual para um determinado nível do fator, segundo Khuri e Cornell (1996) é dado pela diferença entre o mesmo e o ponto médio do intervalo, dividida pela semi-amplitude do intervalo. Nesse caso, a amplitude dos níveis dos fatores codificados pertencem ao intervalo $[-1, 1]$.

Pode-se facilmente notar que conforme o pesquisador aumenta o número de fatores em experimentos fatoriais, o número de tratamentos aumenta rapidamente, podendo inviabilizá-lo. Como exemplo, um delineamento fatorial completo 2^6 tem $2^6 = 64$ tratamentos, sendo que para estimar os parâmetros do modelo com todas as possíveis interações envolvendo os 6 fatores, são necessárias, no mínimo, 64 observações. Segundo Barbin (2003) e Wu e Hamada (2000), $k \geq 7$ já torna o delineamento praticamente inviável. Para fatores com $N = 3$ níveis, o valor de k já é considerado grande se $k \geq 4$. Se, por outro lado, o pesquisador assumir que as interações de maior ordem podem ser desprezíveis, então as informações sobre os efeitos principais e interações de ordem menor podem ser obtidas utilizando apenas uma fração do experimento fatorial completo, chamada fatorial fracionado. Este também é bastante usado quando se considera um grande número de fatores e o pesquisador quer identificar um subconjunto de fatores que têm mais efeito sobre a variável resposta. Considere um delineamento fatorial 2^4 completo, com os fatores A , B , C e D , contendo $2^4 = 16$ tratamentos. Se for necessário usar um número menor de tratamentos, é comum usar metade deste fatorial ($\frac{1}{2}2^4 = 2^{4-1}$) que pode ser obtida a partir de um delineamento fatorial completo 2^3 , como o da tabela 1, adicionando o fator D cujo coeficientes dos contrastes são iguais aos da interação tripla, ou seja, $D = ABC$. Neste delineamento, no entanto, os efeitos principais estão confundidos com as interações triplas e as interações de segunda ordem ficam confundidas entre elas, sendo impossível diferenciar os efeitos confundidos. Assim, no caso das interações duplas, não é possível diferenciar AB de CD , AC de BD e AD de BC . Como consequência, quando se estimam os efeitos de algumas das interações duplas, na realidade são estimados a soma dos efeitos confundidos. No caso, quando se estima o efeito da interação dupla AB , na realidade estima-se $AB + CD$.

Outras frações alternativas, denominadas frações ótimas por Wu e Hamada

(2000), podem ser obtidas por uma metodologia apresentada pelos autores, baseada no conceito de resolução de um delineamento e no critério do mínimo desvio.

Em situações em que as unidades experimentais têm custos muitos elevados, no entanto, não é raro o desejo de se ajustarem modelos com mais parâmetros que observações, ficando inviável realizar a regressão por meio de métodos clássicos. Entretanto, pode-se utilizar uma aproximação Bayesiana seguindo a estratégia de Chipman, Hamada e Wu (1997), que é uma combinação de seleção de variáveis do algoritmo de George e McCulloch (1993), com prioris para relacionar preditores dadas por Chipman (1996). Neste caso, o número de modelos possíveis é grande e podem existir diferentes modelos que explicam os dados observados.

A estratégia de Chipman, Hamada e Wu (1997) utiliza, como priori, alguns princípios clássicos da metodologia de superfície de resposta para ajustar um modelo e estabelece a seleção de modelos por meio das probabilidades a posteriori dos mesmos.

Os três princípios utilizados são o dos efeitos esparsos, o de ordem hierárquica e efeito da hereditariedade. Segundo o efeitos esparsos, o número de importantes efeitos é pequeno em relação ao número total de efeitos. Segundo o princípio da ordem hierárquica, descrito por Wu e Hamada (2000): os efeitos de baixa ordem são mais prováveis de serem importantes do que efeitos de alta ordem e os efeitos de mesma ordem são igualmente prováveis de serem importantes. Segundo os autores, este é um princípio empírico cujo validamento tem sido comprovado por meio de muitos experimentos reais segundo os autores.

O princípio do efeito de hereditariedade, por sua vez, também apresentado pelo autores, estabelece que, caso uma interação entre certos fatores seja significativa, pelo menos um dos fatores formadores dessa interação seja significativo.

Em resumo, estes princípios podem ser considerados na formulação de uma priori para os parâmetros do modelo adotado, de tal forma que os mesmos possam ser estimados, até mesmo no caso de modelos supersaturados, por meio da inferência Bayesiana.

Seja θ o parâmetro ou vetor de parâmetros do modelo de interesse e y uma observação aleatória da variável resposta de interesse.

O teorema de Bayes, base da inferência Bayesiana, estabelece, então, a relação entre a distribuição a posteriori de θ dado y , $p(\theta|y)$, e a distribuição a priori de θ , $p(\theta)$, dada por:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}, \quad (1)$$

em que

$$p(y) = \int_{\theta} p(\theta, y) d\theta = \int_{\theta} p(y|\theta)p(\theta) \quad (2)$$

e $p(\theta, y)$ é a distribuição conjunta de (θ, y) . Para um valor fixo de y , a função $f(y|\theta) = p(y|\theta)$ estabelece, por sua vez, a verossimilhança para cada um dos possíveis valores de θ .

Deve-se salientar ainda, que, neste caso, $p(y)$ dado pela equação (2) é uma constante chamada constante normalizadora. Com isso, o teorema de Bayes, dado pela equação (1), tem sua forma usual dada por

$$p(\theta|y) \propto f(y|\theta)p(\theta). \quad (3)$$

A distribuição a priori tem um importante papel na análise Bayesiana, sendo usada para descrever uma informação sobre os parâmetros desconhecidos antes da obtenção dos resultados do experimento (BOX; TIAO, 1992), ou seja, reflete o conhecimento prévio que o pesquisador tem a respeito de seu problema.

Após definida a distribuição a priori, obtém-se a distribuição a posteriori dos parâmetros, por meio da equação (3). É possível, então, resumir as informações sobre os parâmetros por meio de técnicas inferenciais, utilizando probabilidades e esperanças, que, dependendo da complexidade do problema, não são resolvidas analiticamente. É nesta etapa que se faz necessária a utilização de métodos computacionais intensivos para obter os parâmetros da distribuição a posteriori, como o método de Monte Carlo via Cadeias de Markov, denominado de MCMC. No contexto de técnica Bayesiana, o amostrador de Gibbs e o algoritmo Metropolis-Hastings, descritos por Gelman et al. (1995), são os métodos mais utilizados. Resumidamente, a idéia destes métodos é que a distribuição dos parâmetros obtida por meio dos mesmos converge para a distribuição original (PAULINO; TURKMAN; MURTEIRA, 2003). Estes métodos intensivos estão implementados no pacote computacional R (2007) cuja utilização é descrita detalhadamente por Albert (2007).

Para monitorar a convergência das cadeias geradas pelo métodos citados, podem ser utilizados, então, os métodos diagnósticos propostos por Raftery e Lewis (1992) e Geweke

(1992). O método de Geweke (1992) se baseia no teste de igualdade de médias da primeira e última parte da cadeia, geralmente dos primeiros 10% e dos últimos 50% elementos da mesma, respectivamente. Neste teste, a cadeia converge quando o valor Z calculado esta entre os limites de uma distribuição normal padronizada, ou através da estimativa do valor p calculado, em que, caso esse valor for menor que um nível de significância adotado pelo pesquisador, conclui-se que a cadeia não convergiu.

O método de Raftery e Lewis (1992), por sua vez, baseia-se na acurácia de estimação do quantil das cadeias geradas. Para isso, considere a distância mínima de uma iteração a outra (*thin*) para se obter uma subamostra aproximadamente independente e também o número de iterações iniciais que devem ser descartadas (*burn-in*). A convergência é avaliada, então, por meio do chamado fator de dependência, dado pelo acréscimo multiplicativo ao número de iterações necessárias para se alcançar a convergência. Assim, segundo os autores, a cadeia não converge enquanto esse fator for maior que cinco. Esses testes de convergência foram implementados, usando-se o pacote computacional *R*, no pacote de funções *boaSmith* (2009).

Uma vez obtida a convergência das cadeias, obtêm-se as probabilidades a posteriori de diversos modelos alternativos pertencentes a uma mesma família de modelos.

Finalmente, para analisar a diversidade de modelos a posteriori pode-se, então, utilizar o índice de Shannon (1948), dado por:

$$H = -\ln \left(\prod_{l=1}^k \gamma_l^{\gamma_l} \right) = -\sum_{l=1}^k \gamma_l \ln(\gamma_l) = -\sum_{l=1}^k \gamma_l \ln(\gamma_l), \quad (4)$$

sendo k o número de modelos a posteriori e γ_l a probabilidade do l -ésimo modelo a posteriori. O valor de H é igual a 0 se e somente se houver um modelo a posteriori que tenha probabilidade γ_l igual a 1, ou seja, se houver apenas um modelo a posteriori. Por outro lado, o valor máximo para H , para um determinado k , é $\ln(k)$, ocorrendo quando $\gamma_l = \frac{1}{k}$ ($l = 1, \dots, k$), ou seja, quando todos os modelos tem iguais probabilidades a posteriori de serem escolhidos. Resumidamente, quanto maior a diversidade dos modelos, maior é o índice H .

3 MATERIAIS E MÉTODOS

3.1 Materiais

Como materiais, são utilizados neste trabalho dois conjuntos de dados, o primeiro obtido por meio de simulação e o segundo, por meio de um experimento.

O primeiro considera três fatores A , B e C ou variáveis explicativas x_1 , x_2 e x_3 , cada uma com dois níveis -1 ou 1 no esquema fatorial 2^3 , conforme apresentado na tabela 3, cujos resultados, foram obtidos por meio de simulação considerando-se o modelo:

$$y_i = 10 + 5x_{1i} + 7x_{2i} + 5x_{1i}x_{2i} + 2x_{2i}x_{3i} + \epsilon_i, \quad (5)$$

em que $\epsilon_i \sim N(0, 4)$, $i = 1, \dots, 8$, e $Cov(\epsilon_i, \epsilon_{i'}) = 0$, para todo $i \neq i'$.

Tabela 3 - Níveis dos fatores do delineamento 2^3 completo, e observações da variável resposta y , obtidas por meio de simulação

	A	B	C	
Tratamento	x_1	x_2	x_3	y
1	—	—	—	5,04
2	—	—	+	0,63
3	—	+	—	2,25
4	—	+	+	7,80
5	+	—	—	5,59
6	+	—	+	1,78
7	+	+	—	22,58
8	+	+	+	28,27

Como segundo conjunto de dados, considerou-se os resultantes de um experimento para a formulação de biofilmes de amido de ervilha (*Pisum sativum*) conduzido por Manoel Divino da Matta Junior, no Departamento de Ciência e Tecnologia de Alimentos da ESALQ-USP, em Piracicaba, São Paulo, em 2009. Neste, as diferentes formulações utilizadas constavam de: 3, 4 ou 5% de amido; 0, 0,05 ou 0,1% de goma e 0, 10 ou 20% de glicerol em

combinações dadas seguindo o esquema fatorial 2^3 com 5 repetições do ponto central em um delineamento inteiramente ao acaso. Para o presente trabalho, selecionou-se dentre várias, a variável resposta alongamento sob força máxima em milímetros. As variáveis explanatórias X_1 = porcentagem de amido, X_2 = porcentagem de goma e X_3 = porcentagem de glicerol foram, então, codificadas por meio de $x_1 = \frac{2X_1-8}{2}$, $x_2 = \frac{2X_2-0,1}{0,1}$ e $x_3 = \frac{2X_3-20}{20}$ e apresentadas na tabela 4. Note, nesta tabela, que não houve formação de biofilme com a formulação 5, que considera $(x_1, x_2, x_3) = (1, -1, -1)$, ou seja, $X_1 = 5\%$ de amido, $X_2 = 0\%$ de goma e $X_3 = 0\%$ de glicerol. Note, ainda, que a formulação 9, que considera as porcentagens intermediárias de amido, goma e glicerol, foi repetida cinco vezes.

Tabela 4 - Formulações de diferentes biofilmes de ervilha, coeficientes para estimação das interações e alongamentos sob força máxima (y) dos biofilmes formados

i	Formulação	Amido	Goma	Glicerol	Alongamento			
	(Tratamento)	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	y
1	1	-1	-1	-1	1	1	1	4,7458
2	2	-1	-1	1	1	-1	-1	11,1002
3	3	-1	1	-1	-1	1	-1	5,6792
4	4	-1	1	1	-1	-1	1	10,7002
5	5	1	-1	-1	-1	-1	1	—
6	6	1	-1	1	-1	1	-1	12,7502
7	7	1	1	-1	1	-1	-1	7,1750
8	8	1	1	1	1	1	1	25,8150
9	9	0	0	0	0	0	0	4,6190
10	9	0	0	0	0	0	0	3,8922
11	9	0	0	0	0	0	0	5,9042
12	9	0	0	0	0	0	0	4,9750
$n = 13$	9	0	0	0	0	0	0	5,5378

3.2 Métodos

Considere o modelo linear geral, dado por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (6)$$

em que $\mathbf{y}$ é o vetor de observações com dimensão n ; $\mathbf{X}$ é a matriz do delineamento, de dimensões $n \times (p + 1)$; $\boldsymbol{\beta}$ é o vetor com dimensão $p + 1$, incluindo o intercepto e demais p parâmetros correspondentes aos efeitos fatoriais e $\boldsymbol{\epsilon}$ é o vetor de erros aleatórios com distribuição normal multivariada $MN(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, em que $\mathbf{0}$ é o vetor de zeros, com dimensão n e $\mathbf{I}_n = \text{diag}\{1, 1, \dots, 1\}$ é a matriz identidade.

O princípio de hereditariedade pode ser, então, incorporado diretamente à priori considerando-se, para isso, um vetor $\boldsymbol{\delta}$ de dimensão $p + 1$, formado por 0's e 1's, cujo j -ésimo elemento, δ_j , é 0, se β_j , ($j = 1, 2, \dots, p + 1$) não é importante, ou 1 se β_j é importante. Considera-se, ainda, que $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\delta}^T, \sigma^2)^T$ é o vetor de parâmetros a serem estimados, com dimensão $2p + 3$. Para completar a formulação Bayesiana, são apresentadas, na sequência, as distribuições a priori para $\boldsymbol{\theta}$. Considere, inicialmente que a distribuição a priori para o j -ésimo elemento de $\boldsymbol{\beta}$ seja a mistura de normais dada por

$$\pi(\beta_j | \delta_j) = \begin{cases} N(0, \sigma^2 \tau_j^2) & \text{se } \delta_j = 0 \\ N(0, \sigma^2 (c_j \tau_j)^2) & \text{se } \delta_j = 1 \end{cases} \quad (7)$$

Neste caso, os hiperparâmetros c_j e τ_j devem ser especificados de maneira tal que, quando $\delta_j = 0$, β_j tenha um valor próximo a zero e quando $\delta_j = 1$, exista uma probabilidade elevada de $|\beta_j|$ ter um valor alto.

A distribuição a priori para σ^2 considerada é a gama invertida,

$$\sigma^2 \sim IG(\nu/2, \nu\lambda/2) \quad (8)$$

sendo ν e λ seus hiperparâmetros, cuja escolha será abordada posteriormente.

Finalmente, como $\boldsymbol{\delta}$ especifica um determinado modelo, a priori para $\boldsymbol{\delta}$ especifica a priori para cada modelo possível e é dada por:

$$\pi(\boldsymbol{\delta}) = \prod_{j=1}^{p+1} p_j^{\delta_j} (1 - p_j)^{1-\delta_j}, \quad (9)$$

em que $p_j = \text{Prob}(\delta_j = 1)$, considerando independências entre as prioris para δ_j e $\delta_{j'}$, $j \neq j'$. Esta consideração, no entanto, não é apropriada para modelos com interações. Neste caso, deve-se usar uma priori hierárquica para $\boldsymbol{\delta}$, na qual é incorporado o princípio da hereditariedade.

Considere, para fins de ilustração, que há três fatores, A, B e C, com dois níveis cada e que apenas os efeitos principais e as suas interações duplas devem fazer parte do modelo. Neste caso, tem-se que $\boldsymbol{\delta} = (\delta_{Int}, \delta_A, \delta_B, \delta_C, \delta_{AB}, \delta_{AC}, \delta_{BC})$ e baseando-se no princípio de hereditariedade, a significância de AB depende da significância de A e de B , ou seja, se nenhum dos efeitos principais é significativo no modelo, então a interação parece menos plausível. Esse princípio pode ser incorporado à priori para $\boldsymbol{\delta}$, assumindo que os termos de mesma ordem são independentes. Com isso, os valores de $\boldsymbol{\delta}$ associados aos termos de primeira ordem, δ_A, δ_B e δ_C , são independentes e os associados aos termos de segunda ordem, δ_{AB}, δ_{AC} e δ_{BC} , também a são. Assim, o princípio da hereditariedade assume que o termo com interação depende apenas dos termos que o formam e no caso, $\text{Prob}(\delta_{AB}|\delta_A, \delta_B, \delta_C) = \text{Prob}(\delta_{AB}|\delta_A, \delta_B)$. Logo, tem-se que:

$$\begin{aligned} \text{Prob}(\boldsymbol{\delta}) &= \text{Prob}(\delta_{Int})\text{Prob}(\delta_A)\text{Prob}(\delta_B)\text{Prob}(\delta_C) \\ &\times \text{Prob}(\delta_{AB}|\delta_A, \delta_B)\text{Prob}(\delta_{AC}|\delta_A, \delta_C)\text{Prob}(\delta_{BC}|\delta_B, \delta_C). \end{aligned} \quad (10)$$

Este princípio vai ser definido, então, por meio da forma como as interações dependem dos termos que as formam. No caso em questão, a probabilidade da interação AB ser ativa ou importante, ou seja, de $\delta_{AB} = 1$, dados δ_A e δ_B , pode ser escrita como:

$$\text{Prob}(\delta_{AB} = 1|\delta_A, \delta_B) = \begin{cases} p_{00} & \text{se } (\delta_A, \delta_B) = (0, 0) \\ p_{01} & \text{se } (\delta_A, \delta_B) = (0, 1) \\ p_{10} & \text{se } (\delta_A, \delta_B) = (1, 0) \\ p_{11} & \text{se } (\delta_A, \delta_B) = (1, 1) \end{cases}, \quad (11)$$

em que p_{00}, p_{01}, p_{10} e p_{11} são constantes entre 0 e 1 que refletem a intensidade da hereditariedade.

No caso de se ter fatores quantitativos e modelos de segunda ordem ou superior, esta forma de se especificar a distribuição a priori também pode ser utilizada. Considere, por exemplo, um modelo polinomial de quarta ordem em que A^2B^2 é um termo de quarta ordem.

Este fator, segundo o princípio da hereditariedade, depende dos termos de terceira ordem A^2B e AB^2 conforme ilustra a figura 1. Os termos de terceira ordem, por sua vez, dependem dos termos de segunda ordem A^2 , B^2 e AB que o formam. Finalmente, os termos A^2 e B^2 dependem apenas de um termo, ou seja, de A e B , respectivamente e AB depende de A e B . Assim,

$$Prob(\delta_{A^2} = 1 | \delta_A) = \begin{cases} p_{00} & \text{se } \delta_A = 0 \\ p_{11} & \text{se } \delta_A = 1 \end{cases},$$

chamado de princípio da hereditariedade imediato.

Figura 1 - Esquema de dependência do termo de quarta ordem A^2B^2 segundo o princípio da hereditariedade

Caso haja três fatores principais e o modelo desejado inclua termos de ordem até três, o esquema de dependência destas interações, segundo o princípio da hereditariedade é apresentado na figura 2 e a probabilidade de δ é dada por

$$\begin{aligned} Prob(\delta) &= Prob(\delta_{Int})Prob(\delta_A)Prob(\delta_B)Prob(\delta_C) \\ &\times Prob(\delta_{A^2}|\delta_A)Prob(\delta_{B^2}|\delta_B)Prob(\delta_{C^2}|\delta_C) \\ &\times Prob(\delta_{AB}|\delta_A, \delta_B)Prob(\delta_{AC}|\delta_A, \delta_C)Prob(\delta_{BC}|\delta_B, \delta_C) \\ &\times Prob(\delta_{A^2B}|\delta_{A^2}, \delta_{AB})Prob(\delta_{A^2C}|\delta_{A^2}, \delta_{AC})Prob(\delta_{B^2C}|\delta_{B^2}, \delta_{BC}) \\ &\times Prob(\delta_{AB^2}|\delta_{B^2}, \delta_{AB})Prob(\delta_{AC^2}|\delta_{C^2}, \delta_{AC})Prob(\delta_{BC^2}|\delta_{C^2}, \delta_{BC}) \\ &\times Prob(\delta_{ABC}|\delta_A, \delta_B, \delta_C). \end{aligned} \tag{12}$$

Uma vez definidas as distribuições a priori, a estimação Bayesiana das covariáveis selecionadas é realizada por meio das distribuições condicionais. No caso em questão, a dis-

Figura 2 - Esquema de dependência dos termos de modelos polinomiais contendo três fatores principais e interações de ordem até três, segundo o princípio da hereditariedade

tribuição conjunta é dada por

$$\begin{aligned}
 f(\boldsymbol{\theta}, \mathbf{y}) &= f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta})\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}) \\
 &= f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta})\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta})\pi(\sigma^2|\boldsymbol{\delta})\pi(\boldsymbol{\delta}) \\
 &= f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2)\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta})\pi(\sigma^2)\pi(\boldsymbol{\delta}).
 \end{aligned} \tag{13}$$

A função de verossimilhança, por sua vez, considera que $(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) \sim MN(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$, ou seja, uma distribuição normal multivariada com esperança $\mathbf{X}\boldsymbol{\beta}$ e matriz de variância e covariâncias $\sigma^2\mathbf{I}_n$, que tem a forma

$$f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-n} \exp \left\{ \frac{-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2} \right\}, \tag{14}$$

em que n é o número de observações.

A distribuição a priori para $\boldsymbol{\beta}$, dada em (7), pode ser escrita, alternativamente, como a normal multivariada $(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}) \sim MN(\mathbf{0}, \sigma^2\mathbf{D}_{\boldsymbol{\delta}})$, que tem a forma

$$\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}) \propto \sigma^{-(p+1)} \exp \left\{ \frac{-\boldsymbol{\beta}^T \mathbf{D}_{\boldsymbol{\delta}}^{-1} \boldsymbol{\beta}}{2\sigma^2} \right\}, \tag{15}$$

em que $\mathbf{D}_\delta$ é uma matriz diagonal dada por

$$\mathbf{D}_\delta = \text{diag}\{(c_1^{\delta_1} \tau_1)^2, \dots, (c_{p+1}^{\delta_{p+1}} \tau_{p+1})^2\}.$$

A distribuição a priori para σ^2 , que é uma gama invertida dada em (8), tem a forma

$$\pi(\sigma^2) \propto (\sigma^2)^{-(\frac{\nu}{2}+1)} \exp \left\{ -\frac{\nu\lambda}{2\sigma^2} \right\}. \quad (16)$$

e a distribuição a priori para δ é uma multinomial com 2^{p+1} valores, dada em (9).

A seguir será visto como obter as distribuições condicionais completas para β , σ^2 e δ_j ($j = 1, \dots, p+1$), por meio da equação (13).

A distribuição condicional completa para β é calculada por meio do termos de (13) que envolvem β , ou seja,

$$f(\beta|\sigma^2, \delta, \mathbf{y}) \propto f(\mathbf{y}|\beta, \sigma^2) \pi(\beta|\sigma^2, \delta). \quad (17)$$

Como, a distribuição a priori para β dada por (15) e a função de verossimilhança (14) seguem distribuições normais multivariadas, então a distribuição condicional completa para β também segue uma distribuição normal multivariada dada por:

$$(\beta|\sigma^2, \delta, \mathbf{y}) \sim MN(\sigma^{-2} \mathbf{A}_\delta \mathbf{X}^T \mathbf{y}, \mathbf{A}_\delta), \quad (18)$$

sendo $\mathbf{A}_\delta = \sigma^2(\mathbf{X}^T \mathbf{X} + \mathbf{D}_\delta^{-1})^{-1}$ e tem a forma

$$f(\beta|\sigma^2, \delta, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left[\beta^T (\mathbf{D}_\delta^{-1} + \mathbf{X}^T \mathbf{X}) \beta - \beta^T \mathbf{X}^T \mathbf{y} - \mathbf{y}^T \mathbf{X} \beta \right] \right\}. \quad (19)$$

Analogamente, obtém-se a distribuição condicional completa para σ^2 deixando-se de fora o termo $\pi(\delta)$ da equação (13), ou seja,

$$f(\sigma^2|\beta, \delta, \mathbf{y}) \propto f(\mathbf{y}|\beta, \sigma^2) \pi(\beta|\sigma^2, \delta) \pi(\sigma^2). \quad (20)$$

Simplificando a equação (20) tem-se que a distribuição condicional completa para σ^2 é dada por

$$f(\sigma^2|\beta, \delta, \mathbf{y}) \propto (\sigma^2)^{-(\frac{n+p+1+\nu}{2}+1)} \exp \left\{ -\frac{\nu\lambda + (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \beta^T \mathbf{D}_\delta^{-1} \beta}{2\sigma^2} \right\}.$$

Logo, a distribuição condicional completa para σ^2 é uma gama invertida dada por

$$(\sigma^2|\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{y}) \sim IG\left(\frac{1}{2}(n+p+1+\nu), \frac{1}{2}\left[\nu\lambda + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\beta}^T \mathbf{D}_{\boldsymbol{\delta}}^{-1} \boldsymbol{\beta}\right]\right). \quad (21)$$

Para a distribuição condicional completa para $\boldsymbol{\delta}$, basta tirar de (13) os termos que não dependem de $\boldsymbol{\delta}$, ou seja, $f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2)$ e $\pi(\sigma^2)$. Assim, tem-se que

$$f(\boldsymbol{\delta}|\boldsymbol{\beta}, \sigma^2, \mathbf{y}) \propto \pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta})\pi(\boldsymbol{\delta}). \quad (22)$$

A distribuição condicional completa para cada δ_j é uma Bernoulli com o parâmetro a posteriori $Pr(\delta_j = 1|\boldsymbol{\delta}_{-j}, \boldsymbol{\beta}, \sigma^2, \mathbf{y})$. No entanto, é necessário determinar a probabilidade de $\delta_j = 1$ e a distribuição condicional completa para $\delta_j = 1$ tem a forma

$$f(\delta_j|\boldsymbol{\delta}_{-j}, \boldsymbol{\beta}, \sigma^2, y) \propto \pi(\boldsymbol{\beta}|\sigma^2, \delta_j, \boldsymbol{\delta}_{-j})\pi(\delta_j, \boldsymbol{\delta}_{-j}), \quad (23)$$

em que $\boldsymbol{\delta}_{-j} = (\delta_1, \delta_2, \dots, \delta_{j-1}, \delta_{j+1}, \dots, \delta_{p+1})$. Então,

$$\begin{aligned} Pr(\delta_j = 1|\boldsymbol{\delta}_{-j}, \boldsymbol{\beta}, \sigma^2, \mathbf{y}) &= \\ &= \frac{\pi(\boldsymbol{\beta}|\delta_j = 1, \boldsymbol{\delta}_{-j}, \sigma^2)\pi(\delta_j = 1, \boldsymbol{\delta}_{-j})}{\pi(\boldsymbol{\beta}|\delta_j = 1, \boldsymbol{\delta}_{-j}, \sigma^2)\pi(\delta_j = 1, \boldsymbol{\delta}_{-j}) + \pi(\boldsymbol{\beta}|\delta_j = 0, \boldsymbol{\delta}_{-j}, \sigma^2)\pi(\delta_j = 0, \boldsymbol{\delta}_{-j})} \\ &= \frac{\pi(\delta_j = 1, \boldsymbol{\delta}_{-j})}{\pi(\delta_j = 1, \boldsymbol{\delta}_{-j}) + \pi(\delta_j = 0, \boldsymbol{\delta}_{-j})\pi(\boldsymbol{\beta}|\delta_j = 0, \boldsymbol{\delta}_{-j}, \sigma^2)/\pi(\boldsymbol{\beta}|\delta_j = 1, \boldsymbol{\delta}_{-j}, \sigma^2)}. \end{aligned} \quad (24)$$

Concluídas as distribuições condicionais completas é preciso fazer a estimação do vetor de parâmetros $\boldsymbol{\theta}$.

Nota-se que a distribuição a posteriori dada em (13) não é conhecida. Logo, são utilizadas as distribuições condicionais completas para realizar uma amostragem para cada parâmetro e assim estimá-los. Para isso, utiliza-se a técnica MCMC (*Markov Chain Monte Carlo*), mas como as distribuições condicionais completas são conhecidas pode-se utilizar uma técnica mais simples de MCMC chamada amostrador de Gibbs, descrita por Gelman et al. (1995).

Neste trabalho, $\boldsymbol{\theta} = (\beta_0, \beta_1, \dots, \beta_p, \sigma^2, \delta_1, \delta_2, \dots, \delta_{p+1})^T$ e a implementação do amostrador de Gibbs resulta em uma amostra para cada parâmetro.

O procedimento é iniciado por meio de uma escolha arbitrária $\boldsymbol{\theta}^0$ para o vetor de parâmetros $\boldsymbol{\theta}$, que é renovado obtendo-se $\boldsymbol{\theta}^1$. Para essa renovação, o vetor $\boldsymbol{\beta}$ é renovado automaticamente, mas, no caso do vetor $\boldsymbol{\delta}$, é necessário renovar cada um de seus elementos (δ_j , $j = 1, \dots, p+1$), um de cada vez, conforme ilustrado em (25).

$$\begin{aligned}
\boldsymbol{\beta}^1 &\sim f(\boldsymbol{\beta}^1 | \boldsymbol{\beta}^0, (\sigma^2)^0, \boldsymbol{\delta}^0, \mathbf{y}) \\
(\sigma^2)^1 &\sim f((\sigma^2)^1 | \boldsymbol{\beta}^1, (\sigma^2)^0, \boldsymbol{\delta}^0, \mathbf{y}) \\
\delta_1^1 &\sim f(\delta_1^1 | \boldsymbol{\beta}^1, (\sigma^2)^1, \boldsymbol{\delta}^0, \mathbf{y}) \\
\delta_2^1 &\sim f(\delta_2^1 | \boldsymbol{\beta}^1, (\sigma^2)^1, \delta_1^1, \boldsymbol{\delta}_{(-1)}^0, \mathbf{y}) \\
\delta_3^1 &\sim f(\delta_3^1 | \boldsymbol{\beta}^1, (\sigma^2)^1, \delta_1^1, \delta_2^1, \boldsymbol{\delta}_{(-1,-2)}^0, \mathbf{y}) \\
&\vdots \\
\delta_{p+1}^1 &\sim f(\delta_{p+1}^1 | \boldsymbol{\beta}^1, (\sigma^2)^1, \boldsymbol{\delta}_{(-p-1)}^0, \delta_{p+1}^0, \mathbf{y}).
\end{aligned} \tag{25}$$

Esta renovação, na prática, é feita longa o suficiente ($\boldsymbol{\theta}^m$), até que os parâmetros converjam.

Para implementar este procedimento, no entanto, deve-se especificar os hiperparâmetros τ e c da priori de mistura de normais (7) e (19) para $\boldsymbol{\beta}$, e ν e λ da priori para σ^2 que tem distribuição gama invertida dada em (20). Deve-se, ainda especificar os hiperparâmetros p_{00} , p_{01} , p_{10} e p_{11} adotados em (11) para especificação da priori para $\boldsymbol{\delta}$.

Box e Meyer (1986) aconselham usar $c_j = 10$, pois indica um grande efeito para os β_j dado $\delta_j = 1$.

Para a escolha de τ_j , por sua vez, George e McCulloch (1993), sugerem utilizar $\tau_j = \tau_j^*$, em que τ_j^* é dado por

$$\tau_j^* = \frac{\Delta y}{3\Delta x_j}, \tag{26}$$

($j = 1, \dots, p+1$) em que $\Delta \mathbf{y}$ representa uma pequena mudança na variável resposta $\mathbf{y}$ e Δx_j uma grande mudança em x_j . Isto é feito porque se $\delta_j = 0$, então uma grande mudança em x_j é improvável de acarretar mais do que uma pequena mudança em $\mathbf{y}$. Os autores definem Δx_j e $\Delta \mathbf{y}$ como

$$\Delta x_j = \text{máx}(x_{j1}, x_{j2}, x_{j3}, \dots, x_{jn}) - \text{mín}(x_{j1}, x_{j2}, x_{j3}, \dots, x_{jn}) \tag{27}$$

e

$$\Delta \mathbf{y} = \frac{\sqrt{\text{Var}(y_1, y_2, y_3, \dots, y_n)}}{5}. \quad (28)$$

Deve-se salientar que a escolha de τ pode influenciar diretamente nos resultados, e conseqüentemente, na escolha de um bom modelo. Uma maneira de verificar se há essa influência, é atribuir a τ , três valores diferentes dados por $\frac{\tau^*}{2}$, τ^* e $2\tau^*$ (GEORGE; MCCULLOCH, 1993).

Considere, agora, a escolhas dos hiperparâmetros ν e λ . Quando o número de covariáveis é próximo ao número de observações, a priori para σ resulta em um valor muito pequeno para σ resultando em muitos termos considerados ativos no modelo. Uma maneira de corrigir esta deficiência é escolhendo uma priori informativa para σ . A escolha de ν é frequentemente próxima de 2, indicando uma priori não informativa. Para λ , a escolha depende de experimento para experimento uma vez que depende da escala da variável resposta. Assim uma boa afinação para λ resulta em modelos que não têm muitos e nem poucos termos. Convém salientar que estes hiperparâmetros têm que ser escolhidos e afinados de maneira correta, pois podem influenciar nas conclusões das análises.

Quanto à escolha das constantes que refletem a intensidade da hereditariedade, Chipman, Hamada e Wu (1997), consideram empiricamente que, para uma hereditariedade fraca, p_{00} é bastante pequena, isto é, menor do que $p_{10} = p_{01} = 0,1$ e $p_{11} = 0,25$. Para uma hereditariedade fraca rigorosa, sugerem $p_{00} = 0$ e para a hereditariedade forte a interação somente será significativa se ambos os termos que os formam forem significativos, sugerindo $p_{00} = p_{10} = p_{01} = 0$ e $p_{11} = 0,25$. Resumidamente, tais constantes são apresentadas na tabela 5. A escolhas dessas probabilidades pelos autores foram feitas de modo a considerar que na priori para δ sejam incluídos os princípios apresentados anteriormente, ou seja, de modo que: (i) o número de termos importantes em um experimento fatorial seja pequeno; (ii) os efeitos de ordem menor têm maior chance de serem importantes e (iii) caso um efeito principal seja significativo, as probabilidades de as interações que incluem este efeito principal sejam significativas aumenta.

Assim as distribuições condicionais completas e seus hiperparâmetros estão concluídos.

Convém ressaltar que ao optar pela hereditariedade forte, $p_{10} = p_{01} = 0$, e a

priori $\pi(\delta_j = 0, \boldsymbol{\delta}_{-j})$ será, conseqüentemente, igual a 0 para os efeitos principais. Com isso o denominador da equação (24) fica igual ao numerador, fazendo com que a probabilidade dos efeitos principais sejam iguais a 1. Assim, quando se utiliza hereditariedade forte a distribuição condicional completa para δ_j apresentada na equação (24) tem, para todos os fatores principais, probabilidades iguais a 1, e os fatores principais sempre serão considerados no modelo. No entanto, segundo o princípio da hereditariedade, caso uma interação seja significativa, ao menos um dos efeitos principais deve ser significativo e não que todos os efeitos principais sejam necessariamente significativos ou que todos os efeitos principais devam estar no modelo. De modo a corrigir esse problema na especificação da priori, propõem-se para a hereditariedade forte, denominada nova hereditariedade forte, utilizar, $p_{01} = p_{10} = 0,05$, ou seja,

$$Prob(\delta_{AB}|\delta_A, \delta_B) = \begin{cases} 0 & \text{se } \delta_A = \delta_B = 0 \\ 0,05 & \text{se } \delta_A \neq \delta_B \\ 0,25 & \text{se } \delta_A = \delta_B = 1 \end{cases} . \quad (29)$$

Tabela 5 - Sugestões de valores para os hiperparâmetros p_{00} , p_{01} , p_{10} e p_{11} da distribuição a priori de $\boldsymbol{\delta}$ propostos por Chipman, Hamada e Wu (1997) e novos valores propostos neste trabalho (*) para uma hereditariedade forte

Hereditariedade	$p_1 = \dots = p_{p+1}$	p_{00}	$p_{01} = p_{10}$	p_{11}
Fraca	0,25	$0 < p_{00} < 0,1$	0,1	0,25
Fraca-rigorosa	0,25	0	0,1	0,25
Forte(*)	0,25	0	0,05	0,25
Forte	0,25	0	0	0,25

Afim de evidenciar a importância da escolha dos hiperparâmetros da priori para $\boldsymbol{\delta}$, considere o modelo polinomial com $k = 3$ fatores e interações duplas, dado por

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + e. \quad (30)$$

A tabela 6 apresenta a probabilidade que cada termo do modelo vai ter para as diferentes hereditariedade usadas, o número de modelos possíveis e o índice de diversidade H

Tabela 6 - Probabilidades a priori de cada coeficiente do modelo 30 ser importante de acordo com a intensidade da hereditariedade, cujos hiperparâmetros são apresentados na Tabela 5; Números k de possíveis modelos e índice de diversidade H dos possíveis modelos

Coeficiente	Hereditariedade			
	Fraca	Fra. Rigorosa	Forte*	Forte
β_0	0,250	0,250	0,250	0,250
β_1	0,250	0,250	0,250	0,250
β_2	0,250	0,250	0,250	0,250
β_3	0,250	0,250	0,250	0,250
β_{12}	0,059	0,053	0,034	0,016
β_{13}	0,059	0,053	0,034	0,016
β_{23}	0,059	0,053	0,034	0,016
k	128	90	90	36
H	2,815	2,720	2,587	2,355

Figura 3 - Probabilidades a priori dos 14 modelos mais prováveis de acordo com a intensidade da hereditariedade

e na figura 3 são apresentados os 14 modelos mais prováveis para as diferentes probabilidades usadas.

Na tabela 6 pode-se notar que as escolhas dos valores para os hiperparâmetros da priori δ apresentadas na tabela 5 fazem com que as probabilidade de os parâmetros (ou efeitos) serem importantes ou ativos seja sempre menor ou igual a 0,25 expressando deste modo, o princípio dos efeitos esparsos. Nota-se, ainda que os efeitos principais têm sempre probabilidades superiores aos dos efeitos de interações duplas, expressando, por sua vez, o princípio da hierarquia. Finalmente, por meio da comparação dos gráficos da figura 3, nota-se que, com o aumento da intensidade da hereditariedade, a presença de efeitos de interação dupla traz, como consequência a presença de seus efeitos principais formadores, que é a essência do princípio da hereditariedade.

Uma vez especificada as prioris e obtidas as amostras para cada parâmetro é necessário obter os valores centrais das distribuições de cada um deles. Para os parâmetros de β utilizou-se a média e para σ , utilizou-se a mediana, uma vez que sua distribuição condicional completa é uma gama invertida. No entanto, com relação à δ_j , será usado o vetor δ que indica cada modelo, assim, a moda dos vetores é que indica qual dos modelos deve-se usar.

4 RESULTADOS E DISCUSSÃO

Tanto para a análise dos dados da tabela 3 como para os dados de biofilme de amido de ervilha apresentados na tabela 4, foram utilizados os hiperparâmetros $c = 10$, $\nu = 2, 2$, $\lambda = 0,015$ e os demais apresentados na tabela 5, tomando-se, para a hereditariedade fraca, $p_{00} = 0,01$.

Para o primeiro conjunto de dados, $\Delta x_j = 2$ e $\Delta y = \frac{\sqrt{107,4115}}{5} = 2,0728$. Assim, utilizando-se a equação (26), chega-se a $\tau_j^* = \tau^* = 0,3454$ ($j = 1, 2, 3$).

Para o segundo conjunto de dados, $\Delta x_j = 2$ e $\Delta y = \frac{\sqrt{37,9498}}{5} = 1,2321$ e, analogamente, chega-se a $\tau_j^* = \tau^* = 0,2053$ ($j = 1, 2, 3$).

Para avaliar a influência de τ sobre a inferência de $\boldsymbol{\delta}$, escolheu-se segundo a orientação de Chipman, Hamada e Wu (1997), $\tau = \frac{\tau^*}{2}$, $\tau = \tau^*$ e $\tau = 2\tau^*$.

O algoritmo de Gibbs, por sua vez, é iniciado com todos os termos ativos, ou seja, o vetor $\boldsymbol{\delta}$ com todos os elementos iguais a 1, sendo geradas 50000 amostras de $\boldsymbol{\theta}$, descartando as 1000 primeiras amostras (“*burn-in*” = 1000) e armazenam-se somente uma a cada dez amostras sequencialmente (“*thin*” = 10).

4.1 Dados Simulados

A figura 4 apresenta os gráficos dos índices de diversidade (H) dos modelos a posteriori em função do valor do hiperparâmetro τ , segundo a intensidade da hereditariedade. Nesta figura, nota-se claramente que o aumento do valor de τ , assim como o aumento da intensidade da hereditariedade, acarretaram, neste caso, uma diminuição na diversidade dos modelos. Essa diminuição fica mais evidente quando se altera o valor de τ de $\frac{\tau^*}{2}$ para τ^* ou quando se passa da hereditariedade forte proposta neste trabalho para a forte proposta por Chipman, Hamada e Wu (1997). A diminuição acarretada pelo aumento da intensidade da hereditariedade já era esperada uma vez que esse aumento implica na eliminação de diversos modelos.

A tabela 7 exhibe os modelos mais prováveis a partir da amostragem da posteriori de $\boldsymbol{\delta}$ e suas respectivas probabilidades, os números de modelos a posteriori e os índices H de diversidade de Shannon. Nesta, pode-se notar que o número de diferentes modelos a posteriori permanece igual a 16, mesmo com a alteração do hiperparâmetro τ , considerando-se a here-

Figura 4 - Índice H de diversidade dos modelos a posteriori segundo o valor do hiperparâmetro τ e da intensidade da hereditariedade: fraca, fraca rigorosa, forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*)

ditariedade forte. No entanto, o aumento de τ acarreta uma maior concentração em certos modelos que fica evidenciada pelo menor valor do índice de diversidade de Shannon.

Observando-se os modelos mais prováveis, cujas probabilidades a posteriori são apresentadas em negrito na tabela 7 e representado na figura 5, nota-se que para $\tau = \frac{\tau^*}{2}$, independente da intensidade da hereditariedade, o modelo mais provável é

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2. \quad (31)$$

Para $\tau = \tau^*$ ou $\tau = 2\tau^*$ e hereditariedades fraca ou fraca rigorosa houve a exclusão de β_3 (ou termo C) e a inclusão de β_{23} (ou termo BC) gerando o modelo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3, \quad (32)$$

que é o modelo de equação (5) a partir da qual foram gerados os dados originais da tabela 3. Note que, no entanto, pelo princípio da hereditariedade, com a inclusão do termo BC (ou do parâmetro β_{23}), além do termo B (ou parâmetro β_2) já presente no modelo, também deveria estar presente o termo C (ou parâmetro β_3). Este fato acaba por acontecer quando se aumenta a intensidade da hereditariedade para forte, gerando o modelo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3. \quad (33)$$

A tabela 7 revela, ainda, que para a hereditariedade forte proposta por Chipman, Hamada e Wu (1997), todos os modelos que não apresentam todos os efeitos principais A,B e C, têm probabilidade nula, comprovando o problema ressaltado na metodologia. Por outro lado, quando se utiliza a hereditariedade forte proposta aqui, tal problema não ocorre.

A tabela 8 e figura 6 apresentam, por sua vez, as probabilidades a posteriori de cada um dos termos do modelo completo ser ativo segundo o valor de τ e a intensidade da hereditariedade. Nesta, as probabilidades a posteriori iguais a 1 para os efeitos principais A, B e C considerando-se a hereditariedade forte de Chipman, Hamada e Wu (1997) revelam que, neste caso, todos os modelos apresentam todos os efeitos principais, problema também evidenciado pela metodologia e solucionada por meio da hereditariedade forte proposta neste trabalho.

Do exposto até aqui, ressalta-se que a adoção de $\tau = \tau^*$, conforme sugerido por George e McCulloch (1993), fez com que as diferentes intensidades de hereditariedade tivessem

Tabela 7 - Probabilidades a posteriori de alguns modelos mais prováveis considerando-se diferentes valores de τ , hereditariedades fraca, fraca rigorosa (fr. rig.) e forte propostas por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*), números de diferentes modelos (k) e índices de diversidade de Shannon (H). Em negrito são ressaltados os modelos mais prováveis para cada valor de τ e cada intensidade de hereditariedade

Modelo	$\tau = \frac{\tau^*}{2}$				$\tau = \tau^*$				$\tau = 2\tau^*$			
	fraca	fr.rig.	forte*	forte	fraca	fr.rig.	forte*	forte	fraca	fr.rig.	forte*	forte
1 - A,B	0,045	0,049	0,020	0	0,058	0,058	0,027	0	0,063	0,076	0,029	0
2 - A,B,C	0,060	0,059	0,105	0,142	0,052	0,057	0,094	0,144	0,048	0,051	0,097	0,126
3 - Int,A,B,C	0,061	0,064	0,106	0,142	0,042	0,041	0,084	0,106	0,031	0,033	0,069	0,094
4 - Int,A,B,AB	0,080	0,078	0,032	0	0,102	0,109	0,051	0	0,100	0,101	0,049	0
5 - A,B,C,AB	0,021	0,024	0,042	0,048	0,013	0,014	0,021	0,032	0,007	0,012	0,015	0,026
6 - Int,A,B,C,AB	0,109	0,128	0,189	0,248	0,104	0,105	0,185	0,244	0,075	0,078	0,147	0,212
7 - Int,A,B,AB,BC	0,043	0,055	0,022	0	0,112	0,124	0,049	0	0,161	0,177	0,084	0
8 - Int,A,B,C,AB,BC	0,079	0,083	0,129	0,150	0,107	0,120	0,225	0,294	0,133	0,145	0,269	0,401
9 - Int,A,B,C,AB,AC,BC	0,018	0,016	0,027	0,035	0,015	0,021	0,029	0,048	0,016	0,016	0,032	0,047
k	100	64	59	16	96	62	58	16	89	56	56	16
H	0,579	0,559	0,520	0,450	0,548	0,531	0,498	0,422	0,537	0,512	0,490	0,401

Figura 5 - Probabilidades a posteriori de alguns modelos mais prováveis considerando-se diferentes valores de τ , hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*) , em que os números de 1 a 9 correspondem aos modelos descritos na tabela 7

Tabela 8 - Probabilidades a posteriori dos termos considerados no modelo serem ativos, considerando-se diferentes valores de τ e hereditariedades fraca, fraca rigorosa (fr.rig.) e forte propostas por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*)

Termo	Parâmetro	$\tau = \frac{\tau^*}{2}$				$\tau = \tau^*$				$\tau = 2\tau^*$			
		fraca	fr.rig.	forte*	forte	fraca	fr.rig.	forte*	forte	fraca	fr.rig.	forte*	forte
Intercepto	β_0	0,649	0,643	0,696	0,696	0,663	0,679	0,717	0,780	0,672	0,690	0,764	0,802
A	β_1	0,880	0,904	0,958	1	0,884	0,914	0,957	1	0,879	0,912	0,962	1
B	β_2	0,905	0,917	0,970	1	0,900	0,929	0,969	1	0,897	0,930	0,965	1
C	β_3	0,634	0,674	0,860	1	0,549	0,556	0,805	1	0,498	0,518	0,769	1
AB	β_{12}	0,517	0,510	0,555	0,563	0,565	0,579	0,618	0,780	0,582	0,610	0,669	0,714
AC	β_{13}	0,184	0,185	0,181	0,184	0,125	0,124	0,116	0,127	0,100	0,101	0,096	0,099
BC	β_{23}	0,284	0,285	0,298	0,303	0,343	0,341	0,361	0,396	0,384	0,415	0,452	0,492

Figura 6 - Probabilidades a posteriori dos termos considerados no modelo serem ativos, considerando-se diferentes valores de τ e hereditariedade fraca, fraca rigorosa (fr.rig.) e forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*)

efeito sobre a escolha do modelos sem haver uma diminuição considerável da diversidade dos possíveis modelos, como desejado.

Uma vez adotado o valor de $\tau = \tau^*$, é apresentada, a seguir, a estimação dos parâmetros considerando-se as diferentes intensidades da hereditariedade.

A figura 7 mostra as densidades não paramétricas dos valores simulados a posteriori de β e σ , em que foram descartadas as mil primeiras iterações (*burn-in* = 1000) armazenando-se em seguida, os valores correspondentes às iterações de dez em dez (*thin* = 10). Nota-se nesta figura que as densidades dos parâmetros quase não se alteram em função das diferentes intensidades da hereditariedade. Além disso, os intervalos de 90% de credibilidade incluem o valor zero para os parâmetros β_3 , β_{13} e β_{23} e não incluem o zero para os parâmetros β_0 , β_1 , β_2 e β_{12} , o que poderia sugerir o modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$. A inclusão eventual de alguns dos demais parâmetros no modelos final se faz no entanto, devido à utilização do princípio da hereditariedade.

Na tabela 9 são apresentados as estimativas Bayesianas dadas pela média dos valores a posteriori para β_j e pela mediana dos valores de σ .

Tabela 9 - Estimativas dos parâmetros do modelo considerando-se $\tau = \tau^*$ e hereditariedades fraca, fraca rigorosa (fr. rig.) e forte propostas por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*)

Termo	Parâmetro	Hereditariedade			
		fraca	fr. rig.	forte*	forte
Intercepto	β_0	7,586	7,689	7,849	8,100
A	β_1	4,922	5,063	5,173	5,261
B	β_2	5,685	5,713	5,822	5,885
C	β_3	0,316	0,249	0,369	0,368
AB	β_{12}	3,787	3,812	3,854	3,989
AC	β_{13}	-0,001	0,056	0,061	0,082
BC	β_{23}	1,636	1,596	1,623	1,665
	σ	4,853	4,779	4,178	3,682

Figura 7 - Densidades a posteriori dos parâmetros, para análise do dados simulados, considerando-se as hereditariedades fraca, fraca rigorosa e forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho e o intervalo de 90% de credibilidade (HPD)

Na figura 8 são apresentadas as cadeias de valores simulados a posteriori já com o *burn-in* e *thin* aplicados, que revelam as aparentes convergências das séries comprovadas por meio dos procedimentos de Geweke (1992) e Raftery e Lewis (1992), cujas estatísticas são apresentadas nas tabelas 10 e 11. Tais conclusões foram obtidas pois, segundo o critério de Geweke, todos os valores da estatística Z pertencem ao intervalo $[-1, 96; 1, 96]$. Segundo o critério de Raftery e Lewis (1992), por sua vez, todos os fatores de dependência (FD) estão menores que cinco.

Tabela 10 - Estatística (Z) para o teste de convergência de Geweke (1992) considerando-se $\tau = \tau^*$ e a hereditariedade fraca, fraca rigorosa (fr. rig.) e forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*),

Termo	Parâmetro	Hereditariedade			
		fraca	fr. rig.	forte*	forte
Intercepto	β_0	-0,89	0,35	-0,43	-0,73
A	β_1	-0,70	-0,67	-0,50	0,82
B	β_2	0,28	-0,46	-0,95	0,98
C	β_3	0,49	1,06	0,33	0,04
AB	β_{12}	-0,61	0,11	1,12	-0,35
AC	β_{13}	-0,90	-1,89	-1,11	1,12
BC	β_{23}	-0,94	-0,93	0,71	-0,36
	σ	0,84	0,59	0,71	1,47

4.2 Dados de biofilme

Na figura 9 é apresentado o gráfico dos índices de diversidade (H) dos modelos a posteriori em função do valor do hiperparâmetro τ , segundo a intensidade da hereditariedade. Nesta figura ve-se que, independente da intensidade da hereditariedade, a utilização de $\tau = \tau^*$ leva a uma menor diversidade dos modelos a posteriori. Além disso, o aumento da intensidade da hereditariedade leva a uma menor diversidade de modelos a posteriori como era o esperado pois esse aumento acarreta a eliminação de certos modelos como fica também evidenciado da

Figura 8 - Cadeias de parâmetros, para a análise dos dados simulados, considerando-se as hereditariedades fraca, fraca rigorosa e forte proposta por Chipman, Hamada e Wu(1997) e forte proposta neste trabalho

Tabela 11 - Estatísticas (FD) para o teste de convergência de Raftery e Lewis (1992) considerando-se $\tau = \tau^*$ e hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta por Chipman, Hamada e Wu (1997) e forte proposta neste trabalho (forte*)

Termos	Parâmetros	Hereditariedade			
		fraca	fr. rig.	forte*	forte
Intercepto	β_0	1,00	1,09	2,21	2,34
A	β_1	1,00	2,05	1,02	1,04
B	β_2	1,07	1,02	1,07	1,11
C	β_3	1,09	1,15	1,13	1,07
AB	β_{12}	1,02	1,15	1,09	1,06
AC	β_{13}	1,09	1,13	1,05	1,04
BC	β_{23}	1,05	1,07	1,03	1,09
	σ	1,07	1,13	1,06	1,07

diminuição dos valores de k , apresentados na tabela 12.

A tabela 12 e figura 10 mostra, dentre outras informações, as probabilidades dos modelos a posteriori mais prováveis quando são usados diferentes valores de τ , e diferentes intensidades da hereditariedade. Nesta fica evidente a influência da hereditariedade especialmente quando se considera $\tau = 2\tau^*$, pois, neste caso, para as hereditariedades fraca e fraca rigorosa o parâmetro β_1 (ou termo Am) não está incluído nos modelos com maior probabilidade, o que acontece quando se adota a hereditariedade forte proposta neste trabalho. No entanto, quando se usa $\tau = \tau^*$, o modelo 6 apresentou, para todas as hereditariedades, maior probabilidade, embora, ainda assim, possa se observar a diminuição da probabilidade do modelo 5 e o aumento da probabilidade do modelo 6 com o aumento da intensidade da hereditariedade, o que evidencia a influência de tal princípio.

Concluindo, quando se usa $\tau = \tau^*$, valor recomendado por George e McCulloch (1993), independente da intensidade da hereditariedade, chega-se ao modelo de equação

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3, \quad (34)$$

Figura 9 - Índice H de diversidade dos modelos a posteriori, para os dados de biofilme, segundo o valor do hiperparâmetro τ e a intensidade da hereditariedade: fraca, fraca rigorosa ou forte proposta neste trabalho (forte*)

Tabela 12 - Probabilidades a posteriori de alguns modelos mais prováveis, para o experimento do Biofilme, considerando-se diferentes valores de τ , hereditariedade fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*), números de diferentes modelos (k) e índices de diversidade de Shannon (H), ressaltando-se, em negrito, o modelo mais provável em cada uma das situações

Modelo	$\tau = \frac{\tau^*}{2}$			$\tau = \tau^*$			$\tau = 2\tau^*$		
	fraca	fr.rig.	forte(*)	fraca	fr.rig.	forte(*)	fraca	fr.rig.	forte(*)
1 - Int,Am,Go,Gli	0,118	0,129	0,122	0,104	0,109	0,150	0,088	0,096	0,171
2 - Int,Am,Go,Gli,Am×Gli	0,088	0,092	0,097	0,062	0,061	0,195	0,039	0,048	0,084
3 - Int,Am,Go,Gli,Am×Go	0,087	0,088	0,086	0,055	0,053	0,097	0,042	0,044	0,082
4 - Int,Am,Go,Gli,Go×Gli	0,050	0,053	0,049	0,023	0,026	0,039	0,013	0,012	0,026
5 - Int,Go,Gli,Am×Go,Am×Gli	0,046	0,048	0,052	0,131	0,128	0,049	0,108	0,118	0,055
6 - Int,Am,Go,Gli,Am×Go,Am×Gli	0,131	0,122	0,129	0,186	0,185	0,274	0,104	0,103	0,196
7 - Int,Am,Go,Gli,Am×Go,Am×Gli,Go×Gli	0,048	0,052	0,054	0,057	0,051	0,089	0,026	0,021	0,041
k	82	62	59	81	62	54	97	61	55
H	0,543	0,534	0,531	0,521	0,516	0,472	0,561	0,541	0,506

Figura 10 - Probabilidades a posteriori de alguns modelos mais prováveis, para os dados de biofilme, considerando-se diferentes valores de τ , hereditariiedades fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*), em que os números de 1 a 7 representam os modelos descritos na tabela 12

que inclui todos os efeitos principais formadores dos efeitos de interação dupla.

A tabela 13 e figura 11 apresentam as probabilidades a posteriori de cada termo ser ativo no modelo, para as diferentes hereditariedade e valores de τ . Nestas pode-se ver que o termo Gli (Glicose) apresenta a maior probabilidade de ser ativo independente da hereditariedade ou valor de τ adotados. Como era de se esperar, com o aumento da intensidade da hereditariedade, os termos principais têm suas probabilidades de serem ativos aumentadas, o que é mais facilmente notado para o termo Am (Amido).

Conforme sugerido por George e McCulloch (1993), a estimação dos parâmetros mostrada a seguir, considera as diferentes hereditariedades e apenas para o valor de $\tau = \tau^*$.

Na figura 12 são apresentadas as cadeias dos valores de β e σ simulados para a análise dos dados de biofilme do amido de ervilha. Considerando-se o descarte das primeiras mil amostras (*burn-in*) e valores das iterações tomados de dez em dez (*thin*). Nesta têm-se que todas as séries dos parâmetros, aparentemente, convergiram, o que é comprovado por meio dos teste de Geweke (1992) e Raftery e Lewis (1992) cujos resultados são apresentados na tabelas 14 e 15. Esta conclusão foi obtida pois todos os valores da estatística do teste de Geweke (1992) pertencem ao intervalo $[-1, 96; 1, 96]$ e os relativos à do teste de Raftery e Lewis (1992) são menores do que cinco.

Na figura 13 são apresentados as densidades não paramétricas a posteriori dos parâmetros β e σ e respectivos intervalos de 90% de credibilidade. Como nas densidades a posteriori dos parâmetros para os dados simulados, as densidades pouco se alteraram com a alteração da intensidade da hereditariedade o que se reflete nas pequenas alterações das estimativas Bayesianas dos parâmetros apresentadas na tabela 16

Observa-se, ainda, a pequena magnitude em valor absoluto, das estimativas dos parâmetros β_1 e β_{23} pertencentes aos intervalos de 90% de credibilidade apresetados na figura 13

Mais uma vez, a inclusão ou não de um termo no modelo leva em consideração a intensidade da hereditariedade, como seria desejado.

Como considerações adicionais, tem-se que, mesmo para situações em que há mais parâmetros do que observações, a metodologia poderia ser naturalmente utilizada sem quaisquer alterações.

Tabela 13 - Probabilidades a posteriori dos termos considerados no modelo serem ativos, para os dados de biofilme, considerando-se diferentes valores de τ e hereditariedades fraca, fraca rigorosa (fr.rig.) e forte proposta neste trabalho (forte*)

Termo	Parâmetro	$\tau = \frac{\tau^*}{2}$			$\tau = \tau^*$			$\tau = 2\tau^*$		
		fraca	fr.rig.	forte(*)	fraca	fr.rig.	forte(*)	fraca	fr.rig.	forte(*)
Intercepto	β_0	0,956	0,951	0,957	0,954	0,965	0,963	0,808	0,820	0,831
Am	β_1	0,779	0,793	0,926	0,657	0,657	0,875	0,565	0,580	0,812
Go	β_2	0,859	0,867	0,954	0,875	0,901	0,959	0,823	0,857	0,932
Gli	β_3	0,970	0,971	0,987	0,973	0,981	0,992	0,951	0,965	0,986
Am $\times$ Go	β_{12}	0,485	0,492	0,503	0,553	0,570	0,596	0,425	0,438	0,459
Am $\times$ Gli	β_{13}	0,493	0,477	0,501	0,572	0,569	0,607	0,437	0,448	0,470
Go $\times$ Gli	β_{23}	0,286	0,289	0,285	0,213	0,220	0,214	0,137	0,142	0,137

Figura 11 - Probabilidades a posteriori dos termos considerados no modelo serem ativos, para os dados de biofilme, considerando-se diferentes valores de τ e hereditariedades fraca, fraca rigorosa (fr.rig.) e forte proposta neste trabalho (forte*), descritos na tabela 13

Figura 12 - Cadeias de parâmetros para a análise dos dados de biofilme, considerando-se as hereditariedades fraca, fraca rigorosa e forte proposta neste trabalho

Tabela 14 - Estatística (Z) para o teste de convergência de Geweke (1992), para o experimento do biofilme, considerando-se $\tau = \tau^*$ e a hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*)

Termos	Parâmetros	Hereditariedade		
		fraca	fr. rig.	forte*
Intercepto	β_0	-1,36	0,51	1,69
A	β_1	-0,08	-1,06	0,14
B	β_2	-0,27	0,37	0,36
C	β_3	0,92	-0,32	-1,19
AB	β_{12}	0,56	0,47	-0,13
AC	β_{13}	0,88	0,06	-0,82
BC	β_{23}	-1,31	-0,77	-1,37
	σ	0,66	-1,67	-1,80

Tabela 15 - Estatística (FD) para o teste de convergência de Raftery e Lewis (1992), para os dados de biofilme, considerando-se $\tau = \tau^*$ e as hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*)

Termos	Parâmetros	Hereditariedade		
		fraca	fr. rig.	forte*
Intercepto	β_0	1,93	2,31	2,27
A	β_1	1,00	1,02	1,09
B	β_2	1,05	1,00	0,99
C	β_3	1,00	1,09	1,00
AB	β_{12}	1,02	0,98	1,02
AC	β_{13}	1,02	0,97	1,15
BC	β_{23}	1,00	1,03	1,02
	σ	1,02	1,00	0,98

Figura 13 - Densidades a posteriori dos parâmetros, para a análise dos dados de Biofilme, para as hereditariedades fraca, fraca rigorosa e forte proposta neste trabalho e o intervalo de 90% de credibilidade (HPD)

Tabela 16 - Estimativas dos parâmetros do modelo, para os dados de biofilme, considerando-se $\tau = \tau^*$ e as hereditariedades fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*)

Termos	Parâmetros	Hereditariedade		
		fraca	fr. rig.	forte*
Intercepto	β_0	7,1651	7,2262	7,1824
A	β_1	0,8147	0,8264	0,8863
B	β_2	2,9436	2,9847	3,1816
C	β_3	5,7348	5,7274	5,8995
AB	β_{12}	2,2229	2,2574	2,3762
AC	β_{13}	2,2609	2,2533	2,4494
BC	β_{23}	-0,0519	-0,0394	-0,0932
	σ	4,0968	4,0262	3,8879

5 CONCLUSÕES

O método utilizado apresentou ser muito prático, pois estima os parâmetros ao mesmo tempo que é selecionado o modelos utilizando os princípios da hereditariedade, hierarquia e efeito esparsos. Porém a especificação dos hiperparâmetros da priori para δ , sugerida por Chipman, Hamada e Wu (1997), apresentou problemas quando se considera a hereditariedade forte, e a alteração, para estes hiperparâmetros, proposta neste trabalho corrige os problemas apresentados.

REFERÊNCIAS

- ALBERT, J.; **Bayesian Computation with R**. New York: Springer, 2007. 267 p.
- BARBIN, D. **Planejamento e Análise Estatística de Experimento Agronômicos**. Arapongas: Midas , 2003. 208 p.
- SMITH, B. J. **BOA**: Bayesian Output Analysis Program for MCMC. URL. Disponível em: <<http://www.public-health.uiowa.edu/boa>>. Acesso em: 16 maio 2009.
- BOX, G. E. P.; MEYER, R. D. An Analysis for Unreplicated Fractional Factorials. **Technometrics**, New York, v. 28, n. 1, p. 11-18, 1986.
- BOX, G. E. P.; TIAO, R. D. **Bayesian inference in statistical analysis**. New York: John Wiley, 1992. 588 p.
- CHIPMAN, H. Bayesian Variable Selection with Related Predictors. **Canadian Journal of Statistics**, Vancouver, v. 24, p. 17-36, 1996.
- CHIPMAN, H.; HAMADA, M.; WU, C. F. J. A Bayesian Variable-Selection Approach for Analyzing Designed Experiments with Complex Aliasing. **Technometrics**, New York, v. 39, n. 4, p. 372-381, 1997.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian Data Analysis**. London: Chapman & Hall, 1995. 526 p.
- GEORGE, E. I.; MCCULLOCH, R. E. Variable Selection Via Gibbs Sampling. **Journal of the American Statistical Association**, Boston, v. 88, n. 423, p. 881-889, 1993.
- GEWEKE, J. **Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments**. Oxford: Oxford University Press, p. 625-631, 1992.
- KHURI, A. I.; CORNELL, J. A. **Response Surfaces: designs and analyses**. New York: Marcel Dekker, 1996. 510 p.
- MYERS, R. H.; MONTGOMERY, D. C. **Response Surface Methodology: process and product optimization using designed experiments**. New York: John Wiley & Sons, 1995. 700 p.
- PAULINO, C. D. ; TURKMAN, A. A. ; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 446 p.
- R. Development Core Team. R Foundation for Statistical Computing. **R**: A language and environment for statistical computing, Vienna, Austria. ISBN 3-900051-07-0, URL. Disponível em: <<http://R-project.org>>. Acesso em: 27 jun 2007.
- RAFTERY, A. L. ; LEWIS, S.; How many iterations in the Gibbs sampler?. In BERNARDO, J. M.; BERGERER, J.O.; DAWID, A.P.; SMITH, A.F.(Ed). **Bayesian statistics**. Oxford: Oxford University, 1992. p. 763-774.

SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, New York, v. 27, pp. 379-423, 623-656, 1948.

WU, C. F. J.; HAMADA, M. **Experiments Planning, Analysis, and Parameter Design Optimization**. New York: John Wiley, 2000. 630 p.

ANEXOS

ANEXO - Programa feito no pacote estatístico R

Limpando qualquer termo existente no R, chamando alguns pacotes necessários para o programa e chamando os dados#####

```
rm(list=ls())
require(MASS)
require(mvtnorm)
require(LearnBayes)
require(boa)
source("dados.r")
```

Fazendo as Funções do programa

Gerando a posteriori condicional completa para betas

```
B.draw.f=function(X,B,D,y,sigma2){
  D=diag(1/(((c^delta)*tau)^2))
  media=solve((t(X)%*%X)+D)%*%t(X)%*%y
  sigma22=sigma2*(solve((t(X)%*%X)+D))
  B= mvrnorm(1,media,sigma22)
return(B)
}
```

####gerando a posteriori condicional completa para sigma2

```
sigma2.draw.f=function(n,v,para,lambda,y,X,B,D){
  a=(1/2)*(n+v+1+para)
  D=diag(1/(((c^delta)*tau)^2))
  b=(1/2)*(v*lambda+(t(y-X)%*%B))%*%(y-X)%*%B+t(B)%*%D)%*%B)
  sigma2=1/rgamma(1,shape=a,rate=b)
return(sigma2)
}
```

gerando as probabilidades dos deltas####3

```
delta.draw.f1=function(delta,sigma2){
for(j in 1:para){
delta[j]=1
D=diag(((c^delta)*tau)^2)
Di=diag(1/(((c^delta)*tau)^2))
#prioriB1=dmvnorm(B,zero,sigma2*D)+1e-300
prioriB1=(1/((2*pi)^(para/2)*sqrt(det(sigma2*D))))*exp(-(1/2)*(B-zero)%*%
((1/sigma2)*Di)%*%(B-zero))
```

```

ay=0
for(i in 2:pa)
ay[i]=ay[i-1]+delta[i]*(pa-2-sum(delta[-c(1,i,(pa+1):para]))))
ay[pa]
by=0
for(i in 2:(pa-1))
by[i]=by[i-1]+delta[i]*sum(delta[-c(1:i,(pa+1):para)])
ci=0
for(i in 2:(pa-1))
ci[i]=ci[i-1]+(1-delta[i])*(pa-i-sum(delta[-c(1:i,(pa+1):para]))))
prioriD1=prod(((p[1]^delta[1])*((1-p[1])^(1-delta[1]))):((p[1]^delta[pa])*
((1-p[1])^(1-delta[pa]))))*((p[2]^(by[pa-1]))*(p[3]^(ay[pa]))*(p[4]^ci[pa-1]))

delta[j]=0
D=diag(((c^delta)*tau)^2)
Di=diag(1/(((c^delta)*tau)^2))
#prioriB0=dmvnorm(B,zero,sigma2*D)+1e-300
prioriB0=(1/((2*pi)^(para/2)*sqrt(det(sigma2*D))))*exp(-(1/2)*(B-zero)%*(
(1/sigma2)*Di)%*(B-zero))
ay=0
for(i in 2:pa)
ay[i]=ay[i-1]+delta[i]*(pa-2-sum(delta[-c(1,i,(pa+1):para]))))
ay[pa]
by=0
for(i in 2:(pa-1))
by[i]=by[i-1]+delta[i]*sum(delta[-c(1:i,(pa+1):para)])
ci=0
for(i in 2:(pa-1))
ci[i]=ci[i-1]+(1-delta[i])*(pa-i-sum(delta[-c(1:i,(pa+1):para]))))
prioriD0=prod(((p[1]^delta[1])*((1-p[1])^(1-delta[1]))):((p[1]^delta[pa])*
(1-p[1])^(1-delta[pa]))))*((p[2]^(by[pa-1]))*(p[3]^(ay[pa]))*(p[4]^ci[pa-1]))

pr=prioriD1/(prioriD1+prioriD0*prioriB0/prioriB1)
delta[j]=rbinom(1,1,pr)
}
delta      }

```

Definindo os hiperparâmetros a serem usados

```

n=8
  c=10
  #tau=(sd(y)/5)/3
  ou
  tau=(sd(y)/5)/6
  ou
  #tau=((sd(y)/5)/6)/2
  v=2.2

```

```

lambda=0.015
p=c(0.25,0.25,0.1,0.01)
pa=4
para=7
zero=rep(0,para)
delta=rep(1,para)
B=rep(7,para)
sigma2=1
D=diag((c^delta)*tau)^2

##### Programa Gibbs #####

B.draw=B
sigma2.draw=sigma2
delta.draw=delta

m=50000
for(i in 1:m){
  B=B.draw.f(X,B,D,y,sigma2)
  B.draw=c(B.draw,B)

  sigma2=sigma2.draw.f(n,v,p,lambda,y,X,B,D)
  sigma2.draw=c(sigma2.draw,sigma2)

  delta=delta.draw.f1(delta,sigma2)
  delta.draw=c(delta.draw,delta)
}

##### indicando o descarte inicial das iterações e distância mínima
de uma iteração para outra #####
burnin=1000
thin=10

##### Obtendo as subamostras com thin e burni-in #####
B.draw=matrix(B.draw,nrow=m+1,ncol=para,T)
delta.draw=matrix(delta.draw,nrow=m+1,ncol=para,T)
delta.dr=matrix(nrow=((m-burnin)/(thin)),ncol=para)
for(i in 1:((m-burnin)/(thin))){
  delta.dr[i,]=delta.draw[thin*i+burnin,]}
b.draw=matrix(nrow=((m-burnin)/(thin)),ncol=7)
for(o in 1:((m-burnin)/(thin))){
  b.draw[o,]=B.draw[(thin*o)+burnin,]}
sig2.draw=sigma2.draw[thin+burnin]
for(i in 1:((m-burnin)/(thin))){
  sig2.draw[i]=sigma2.draw[thin*i+burnin]
}

```

Obtendo as frequências a posteriori, o número e o índice de diversidade de Shannon para os modelos

```
colar<-function(aaa){
paste(aaa,collapse="")}
contag<-as.factor(apply(delta.dr,1,colar))
sort(summary(contag),FALSE)/((m-burnin)/thin)

k=length(summary(contag))
k
H=-sum((summary(contag)/m)*log(summary(contag)/m))
H
```

Plotando alguns graficos usados no trabalho

```
a=sum(delta.dr[,1])
b=sum(delta.dr[,2])
c=sum(delta.dr[,3])
d=sum(delta.dr[,4])
e=sum(delta.dr[,5])
f=sum(delta.dr[,6])
g=sum(delta.dr[,7])
#par(mfrow=c(2,4),pty="s")
plot(c(a,b,c,d,e,f,g)/((m-burnin)/thin),type="h",lwd=5,ylim=c(0,0.99))
)
```

```
par(mfrow=c(2,4),pty="s")
plot(density(b.draw[,1]))
abline(v=0,col="red")
plot(density(b.draw[,2]))
abline(v=0,col="red")
plot(density(b.draw[,3]))
abline(v=0,col="red")
plot(density(b.draw[,4]))
abline(v=0,col="red")
plot(density(b.draw[,5]))
abline(v=0,col="red")
plot(density(b.draw[,6]))
abline(v=0,col="red")
plot(density(b.draw[,7]))
abline(v=0,col="red")
plot(density( sqrt(sig2.draw)),xlim=c(0,80))
```

```
par(mfrow=c(2,4),pty="s")
```

```
plot(b.draw[,1],type="l")
abline(h=0,col="red")
plot(b.draw[,2],type="l")
abline(h=0,col="red")
plot(b.draw[,3],type="l")
abline(h=0,col="red")
plot(b.draw[,4],type="l")
abline(h=0,col="red")
plot(b.draw[,5],type="l")
abline(h=0,col="red")
plot(b.draw[,6],type="l")
abline(h=0,col="red")
plot(b.draw[,7],type="l")
abline(h=0,col="red")
plot(sqrt(sig2.draw),type="l")
abline(h=0,col="red")
```